

# Systematic Coarse-Graining in Nucleation Theory

M. Schweizer<sup>1, a)</sup> and L. M. C. Sagis<sup>2, b)</sup><sup>1)</sup> *ETH Zurich, Department of Materials, Polymer Physics, Vladimir-Prelog-Weg 5, 8093 Zurich, Switzerland.*<sup>2)</sup> *Food Physics Group, Wageningen University, Bomenweg 2, 6703 HD Wageningen, The Netherlands and ETH Zurich, Department of Materials, Polymer Physics, Vladimir-Prelog-Weg 5, 8093 Zurich, Switzerland.*

(Dated: 28 April 2015)

In this work we show that the standard method to obtain nucleation rate-predictions with the aid of atomistic Monte-Carlo simulations leads to nucleation rate predictions that deviate 3 – 5 orders of magnitude from the recent brute-force molecular dynamics simulations [J. Diemand, R. Angéilil, K. K. Tanaka, and H. Tanaka, *J. Chem. Phys.* **139**, 074309 (2013)] conducted in the experimental accessible supersaturation regime for Lennard-Jones argon. We argue that this is due to the truncated state space literature mostly relies on, where the number of atoms in a nucleus is considered the only relevant order parameter. We here formulate the nonequilibrium statistical mechanics of nucleation in an extended state space, where the internal energy and momentum of the nuclei is additionally incorporated. We show that the extended model explains the lack in agreement between the molecular dynamics simulations by Diemand et al. and the truncated state space. We demonstrate additional benefits of using the extended state space; in particular, the definition of a nucleus temperature arises very naturally and can be shown without further approximation to obey the fluctuation law of McGraw and Laviolette. In addition, we illustrate that our theory conveniently allows to extend existing theories to richer sets of order parameters.

PACS numbers: Valid PACS appear here

Keywords: coarse-graining, extended state space, multi-scale modeling

## I. INTRODUCTION

Nucleation is the initial step of a phase transformation, a process ubiquitous from the smallest to the largest physical length-scales: the fabrication of new nanoscale material for various future applications such as the design of new microchips requires a control on the morphology of nucleated material<sup>1,2</sup>. On the micrometer scale, nucleation is for instance relevant in the development of new drugs; here, different nucleation conditions can lead to different packing arrangements of the molecular substance, which would be pharmaceutically classified as different drugs. Therefore a precise handle on the nucleation process is deeply required<sup>3</sup>. On the higher end of length scales, research has been conducted<sup>4</sup> to study cloud seeding - the injection of a substance into clouds to induce rainfall - offering the potential to control the weather and the atmosphere. On cosmic length-scales nucleation leads to creation of interstellar dust after super novae. The dust modifies the spectrum of interstellar radiation, and its composition and time-scale of formation are of key interest<sup>5</sup>. In all these phenomena, small nuclei have to overcome an activation barrier to initiate the phase-transformation and to

grow deterministically. The rate at which small nuclei are capable to escape from the initial fluctuation dominated growth regime is the fundamental quantity of interest in nucleation theory. This nucleation rate has been predicted through a variety of methods ranging from brute force molecular dynamics (MD) simulations<sup>6–12</sup> to phenomenological, semi-phenomenological and statistical mechanics approaches<sup>13–19</sup> among others.

MD simulates all involved atoms in the nucleation process by solving their equations of motion and can be used to obtain the exact nucleation rates under various initial conditions of the metastable phase. Since MD conducts computer experiments with systems of nano sized dimensions, up to limits of a few micrometers, these studies favour regimes in which the nucleation rate is large, so that despite of the microscopic system size critically sized nuclei pop up frequently enough and adequate statistics on the obtained data can be ensured. In the prototypical studied nucleation of liquid drops in a metastable gas, MD works in the large supersaturation regime where the metastable phase is very dense. This is in strong contrast to real experiments. Most laboratory experiments<sup>20–23</sup> are performed at low supersaturations and measure nucleation rates less than  $10^{10}\text{cm}^{-3}\text{s}^{-1}$ . For the nucleation of argon the latest experimental setups through the supersonic Nozzle (SSN) experiment<sup>24</sup> achieved to study nucleation rates up to  $10^{17}\text{cm}^{-3}\text{s}^{-1}$ . Only recently, in 2013, large-

<sup>a)</sup> Electronic mail: marco.schweizer@math.ethz.ch<sup>b)</sup> Electronic mail: leonard.sagis@wur.nl

scale MD simulations<sup>6</sup> were capable to operate in the experimentally accessible regime set by the SSN. In Fig. 1 we have reported the studied nucleation rate in MD simulations of argon in some of the literature throughout the past 20 years - unless sophisticated large-scale simulations are exploited, MD would typically study nucleation regimes that differ by 5 – 10 orders of magnitude from the SSN experiment and 12 – 17 orders of magnitude from typical experiments. MD cannot

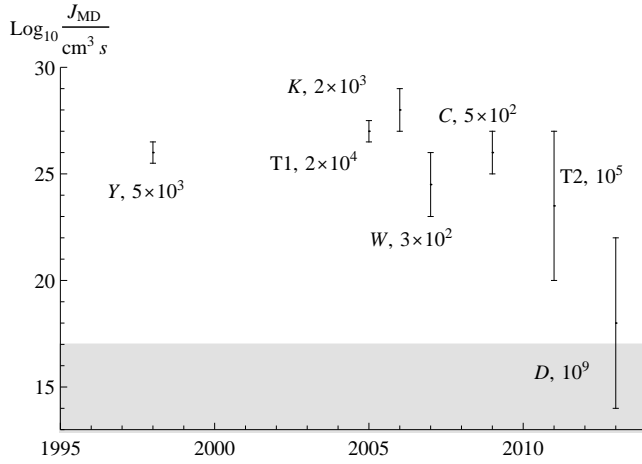


FIG. 1. Illustration of accessible nucleation rate  $J_{MD}$  in selected molecular dynamics studies<sup>6–12</sup> of argon from 1997 – 2013. We marked the first letter of the family name of the first author of the study together with the number of used atoms. The shaded area shows the experimental accessible regime.

be used to study realistic experimental regimes unless tremendous computational effort is made, which is the main drawback of the technique. Coarse-grained models allow to bridge the time- and length scale gap between atomistic dynamics and nucleation as perceived on a coarser scale. Such models offer the possibility to obtain predictions in the experimental accessible nucleation regime and offer hence an appealing alternative to brute-force MD.

The development of coarse-grained models to capture the relevant macroscale physics of nucleation received tremendous amount of interest. These studies mainly focus on obtaining predictions for the shape and height of the nucleation landscape  $V_S(n_c)$  that prohibits small nuclei to reach the critical size  $n_c^*$ . The barrier height  $\Delta V_S$  of  $V_S$  is particularly important and enters the nucleation rate through an exponential factor  $e^{-\Delta V_S}$ . Relatively minor errors in  $\Delta V_S$  are therefore known to give rise to lack in predictions of the nucleation rate of orders of magnitude. In the literature  $k_B T V_S(n_c)$  is known as the Gibbs free energy of formation of a nucleus of

size  $n_c$  in a metastable phase at temperatur  $T$  and pressure  $P$ . A long history of phenomenological, semi-phenomenological and statistical mechanics models have been obtained to predict the work of formation. In particular, the classical nucleation theory<sup>13</sup>, the internally consistent classical nucleation theory<sup>14</sup>, the semiphenomenological model (SP)<sup>15–17</sup>, the revised SP model<sup>18</sup>, the mean-field kinetic theory (MFKT)<sup>13,25</sup>, the extended modified liquid drop dynamic nucleation theory<sup>19</sup> and many more. In our work we aim to obtain quantitative predictions of  $V_S$  from first principle atomistic Monte-Carlo (MC) simulations thereby avoiding the often crude approximations that mesoscopic theories exploit to obtain the nucleation landscape  $V_S$ .

Standard MC simulations<sup>26–32</sup> in the literature exploit small atomistic systems with  $N$  atoms, pressure  $P$  and temperature  $T$  to sample the probability  $\phi_S(n_c)$  of occurrence of a nucleus of size  $n_c$ . The choice of the isobaric-isothermal ensemble in computer simulations mimics the conditions of many real experiments which control exactly these variables to induce the nucleation process. The link of the microscopically obtained probability  $\phi_S(n_c)$  with the macro-physics manifests itself in the relationship  $V_S(n_c) = -\log(\phi_S(n_c))$ . The probability  $\phi_S(n_c)$  is vanishingly small around the critical size  $n_c^*$ , but the MC simulations can be biased through importance sampling techniques such as umbrella-sampling to access regions of the atomistic phase space that contain a nucleus of size  $n_c$ . MC simulations obtain  $\phi_S(n_c)$  even in the experimentally relevant nucleation regime thereby offering a tool that is potentially superior to brute-force MD. In order to obtain nucleation rate predictions  $J$ ,  $\phi_S$  must however enter an adequate evolution equation obtained through supplementary dynamic considerations of the macroscale nucleation process. The Zeldovich equation<sup>13</sup>

$$\frac{\partial f(n_c)}{\partial t} = \frac{\partial}{\partial n_c} D_{n_c, n_c} \left[ \frac{\partial f(n_c)}{\partial n_c} + f(n_c) \frac{\partial V_S(n_c)}{\partial n_c} \right] \quad (1)$$

recognized as thermodynamically admissible<sup>33</sup> is probably the most popular to capture the dynamics of the distribution function  $f(n_c)$  of a nucleus of interest. Here,  $f(n_c)$  denotes the probability that the considered nucleus has size  $n_c$  and  $D_{n_c, n_c}$  is the diffusion coefficient of the diffusion process described by Eq. (1). The Zeldovich equation can be solved approximately in the stationary limit to obtain the nucleation rate as  $J = J_0 e^{-\Delta V_S}$ , where the prefactor  $J_0$  is related to kinetic coefficients and the barrier shape<sup>13</sup>.

We will illustrate that when the nucleation rate of argon droplets in a supersaturated argon gas is obtained through the standard formula

$J = J_0 e^{-\Delta V_S}$ , where  $\Delta V_S$  is constructed from the aforementioned MC technique, a nucleation rate is predicted that deviates 3 – 5 orders of magnitude from exact brute-force MD predictions by Diemand and coworkers<sup>6</sup>. This is particularly severe since the simulations conducted by Diemand et al. consider the practically relevant supersaturation regime. Furthermore, the argon particles are described by the Lennard-Jones model and this is one of the most basic and often used microscopic models to study the nucleation process. The standard approach to obtain the nucleation barrier by MC simulations is therefore incomplete.

We argue that this is an artefact of relying on the fully truncated state space  $x_c^{\text{trc}} = (n_c)$  to describe the nuclei degree of freedom on the coarse-grained scale. Our attempt to improve nucleation rate predictions obtained from MC simulations has its fundament in 1966 in the classical work<sup>34</sup> by Feder et al. who already noticed the need to extend the aforementioned truncated state space in order to account for certain dynamic effects that cannot be captured by the use of a single state variable. In particular, Feder et al. realized that in the nucleation process a significant amount of latent heat is released indicating that the nuclei exhibit rather strong energy fluctuations on top of their size fluctuations. They showed that a proper inclusion of the internal energy  $u_c$  into the nuclei state space leads to corrections to the nucleation rate of 1 – 3 orders of magnitude. The nucleation landscape in their work is therefore extended to two dimensions such that  $V_S = V_S(n_c, u_c)$ . This is confirmed by other work<sup>33,35–37</sup> focusing on nonisothermal corrections. Another argument in favour of equipping the state space with additional degrees of freedom traces back to the “translational-rotational paradox” initiated by Lothe and Pound around 1962<sup>38</sup>. They realized that nuclei are not static objects, but can move and rotate within the metastable phase. In their semi-quantum mechanical treatment they tried to incorporate the translational and rotational contributions into  $V_S(n_c)$  without extending the nuclei state space and ended up with corrections of  $10^{17}$  to the nucleation rate. After incorporating their correction, theoretical predictions of the nucleation rate deviated from experimentally obtained rates so outrageously that the term “translational-rotational paradox” was coined. Many attempts<sup>39–44</sup> tried to resolve the paradox. In this respect the work<sup>44</sup> of Reguera and Rubi presents an elegant solution by incorporating additional degrees of freedom that characterize the nuclei. In particular, they argued that a proper inclusion of the momentum  $p_c$  and angular momentum  $l_c$  in the state space of the

nucleus leads to much smaller corrections of the order of  $10^2 - 10^5$  for typical nucleation regimes. The landscape would hence be extended to  $V_S = V_S(n_c, p_c, l_c)$ .

The idea of extending the state space to capture relevant effects is not new and for certain order parameters, the extended nucleation landscape  $V_S$  has even been studied by statistical mechanics<sup>45–47</sup>. But in this paper we aim to lay the basis for the nonequilibrium statistical mechanics of nucleation with extended state spaces. We here restrict the study to the specific choice  $x_c = (n_c, p_c, u_c)$  and to homogeneous nucleation in a one-component system. We develop a coarse-graining procedure that is reliably predicting nucleation rates in the low nucleation rate regime where critically sized nuclei are sufficiently large so that their dynamics can be captured in the continuum limit through the use of Fokker-Planck equations. This Fokker-Planck equation involves diffusion in  $x_c$ -space and describes the dynamics of  $f(x_c)$ , the probability that at a certain time a specific nucleus has property  $x_c$ . Our coarse-graining procedure relies on the GENERIC (General Equation for Non-Equilibrium Reversible-Irreversible Coupling) projection-operator technique<sup>48,49</sup> and yields atomistic expressions for the building blocks of the aforementioned Fokker-Planck equation. In particular, we obtain a fundamental expression for the nucleation landscape  $V_S(x_c)$  that can easily be sampled in MC simulations and we show that the diffusion coefficients in the Fokker-Planck equation are a simple generalization of expressions already available<sup>50,51</sup> for the fully truncated case  $x_c^{\text{trc}} = (n_c)$ . Through our nonequilibrium statistical mechanics we obtain contributions to the nucleation barrier that have already been found through mesoscopic irreversible thermodynamics arguments concerning momentum corrections and by McGraw and Laviolette for energy fluctuations. In particular, our framework allows to rigorously define a nucleus temperature  $T_c$  that obeys the non-Gaussian behavior derived by McGraw and Laviolette<sup>52</sup> which has been confirmed in direct MD simulations of nucleation<sup>53,54</sup>. We apply our approach to the most recent MD simulation predictions<sup>6</sup> of Diemand et al. and show that the lack of prediction of 3 – 5 orders of magnitude by the truncated theory can be eliminated. On the example of MFKT which uses the truncated state space we illustrate further that we can conveniently extend existing theories to richer state spaces thereby improving the quality of their predictions.

In Sec. II we develop the coarse-graining scheme for nucleation with extended state spaces from microscopic arguments. We show that the

nucleation landscape  $V_S$  is deeply related to the system entropy and that  $V_S$  is the fundamental static building block generating the nucleation dynamics on the coarse-grained scale. We discuss the nucleation landscape and the evolution equation governing  $f(\mathbf{x}_c)$  in some detail and relate it to existing literature results. In Sec. V we apply our results to extend the state space of the MFKT of nucleation and discuss the nucleation of argon droplets in a metastable gas. Finally, in Sec. VI we use our coarse-graining procedure to obtain the nucleation rate in metastable Lennard-Jones gas. We show that the standard coarse-graining procedure leads to erroneous estimates as compared to brute-force MD simulations<sup>6</sup> and that these can be corrected for by our theory.

## II. DETAILED DERIVATION OF COARSE-GRAINING PROCEDURE

### A. Relevant State Variables

We aim to investigate nucleation in a metastable one component system. Prominent examples are nucleation of liquid drops in a gas phase or nucleation of gas bubbles in a liquid phase. We characterize the state  $\mathbf{x}_c = (n_c, \mathbf{p}_c, u_c)$  of the nuclei by means of their number of atoms  $n_c$ , their total momentum  $\mathbf{p}_c$  and their internal energy  $u_c$ . The incorporation of the total momentum allows for Brownian motion of the nuclei and the total energy as an additional variable is necessary to treat non-isothermal effects. The dynamics of a single nucleus is captured by the distribution function  $f_t(\mathbf{x}_c)$  describing the probability that the nucleus has degrees of freedom  $\mathbf{x}_c$  at time  $t$ ; the time-argument will subsequently be suppressed in the notation. The nuclei are assumed to evolve independently, so that the evolution of each nucleus does not affect other nuclei and the single-nucleus distribution function  $f(\mathbf{x}_c)$  is sufficient to capture the relevant macro-scale physics. We hence neglect the break-up of a nucleus into several nuclei or coalescence effects. We further characterize the full system in which nucleation takes place by its total energy  $E_{\text{tot}}$ , total number of particles  $N_{\text{tot}}$ , total momentum  $\mathbf{P}_{\text{tot}}$  and volume  $V_{\text{tot}}$  so that its state is captured by  $\mathbf{x}_{\text{tot}} = (N_{\text{tot}}, \mathbf{P}_{\text{tot}}, E_{\text{tot}}, V_{\text{tot}})$ . Clearly, as  $\mathbf{x}_{\text{tot}}$  consists solely of conserved quantities, the state vector  $\mathbf{x}_{\text{tot}}$  remains constant during the entire evolution. The state variables  $\mathbf{x}_c$  (or respectively  $f(\mathbf{x}_c)$ ) and  $\mathbf{x}_{\text{tot}}$  are merely a splitting of degrees of freedom to capture the essential characteristics of the nucleus and its environment. However,  $\mathbf{x}_{\text{tot}}$  is not the state of the nucleus ambient phase, but involves both, the nucleus and ambient phase degrees of freedom. In Sec. II E we show that upon a suitable transformation, the variables  $\mathbf{x}_{\text{tot}}$  and  $\mathbf{x}_c$  can indeed be used to split the degrees of freedom into nucleus,  $\mathbf{x}_c$ , and environment degrees of freedom,  $\mathbf{x}_{\text{env}}$ . The splitting  $(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$  allows us to carry out the nonequilib-

rium statistical mechanics more straightforward as  $\mathbf{x}_{\text{tot}}$  is a conserved quantity. However, the state of the ambient phase  $\mathbf{x}_{\text{env}}$  is not conserved as it exchanges particles, momentum and energy with the nucleus and as such, the nonequilibrium statistical mechanics is slightly more involved with the splitting  $(\mathbf{x}_{\text{env}}, \mathbf{x}_c)$ .

The macroscopic state variables must be computable by exploiting specific microscopic expressions. On this more detailed level of description, the system is characterized by  $N$  atoms with mass  $m$ , position  $\mathbf{r}_i$  and momentum  $\mathbf{p}_i$  in a box of volume  $V_{\text{tot}}$ . All atoms build the microscopic phase space  $\bar{\Gamma} = (\mathbf{r}_1, \mathbf{p}_1, \dots, \mathbf{r}_N, \mathbf{p}_N)$ . In order to distinguish particles that are part of our nucleus of interest from the remaining ones we introduce the function  $\Pi_{\theta(\bar{\Gamma})}$  on  $\bar{\Gamma}$ . We use the simple position dependent rule  $\Pi_{\theta(\bar{\Gamma})}(\mathbf{r}_i) = 1$  if particle  $i$  is part of the nucleus and 0 otherwise. Notice that in our notation  $\Pi_{\theta(\bar{\Gamma})}(\mathbf{r}_i)$  is seemingly only depending on  $\mathbf{r}_i$ , but this function of course depends on all position degrees of freedom of  $\bar{\Gamma}$ . Popular cluster criterions  $\Pi_{\theta(\bar{\Gamma})}$  of the kind are the Stillinger-cluster definition<sup>55</sup>, and its extension presented by ten Wolde et al.<sup>32,55</sup> (see also Sec. VI B). We obtain the subsequent microscopic definition of the macro-variables  $\mathbf{x}_c$  and  $f(\mathbf{x}_c)$  characterizing the state of the nucleus:

$$\begin{aligned}\Pi_{n_c} &= \sum_i \Pi_{\theta(\bar{\Gamma})}(\mathbf{r}_i), \\ \Pi_{\mathbf{p}_c} &= \sum_i \mathbf{p}_i \Pi_{\theta(\bar{\Gamma})}(\mathbf{r}_i), \\ \Pi_{u_c}^i &= \sum_i \left[ \frac{(\mathbf{p}_i - m \Pi_{\mathbf{v}_c})^2}{2m} + \phi(\mathbf{r}_i) \right] \Pi_{\theta(\bar{\Gamma})}(\mathbf{r}_i), \\ \Pi_{f(\mathbf{x}_c)} &= \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c),\end{aligned}\tag{2}$$

where  $\phi(\mathbf{r}_i)$  denotes the total interaction potential of particle  $i$  due to all other particles. The velocity  $\Pi_{\mathbf{v}_c}$  of the nucleus is obtained from  $\Pi_{n_c} \Pi_{\mathbf{v}_c} = \Pi_{\mathbf{p}_c}$ . By averaging these quantities over phase space with the proper statistical ensemble, we recover the corresponding macrostate variables  $\mathbf{x}_c$ . The variables  $\Pi_{\mathbf{x}_{\text{tot}}} = (\Pi_{N_{\text{tot}}}, \Pi_{\mathbf{P}_{\text{tot}}}, \Pi_{E_{\text{tot}}}, \Pi_{V_{\text{tot}}})$  as the microscopic expressions for  $\mathbf{x}_{\text{tot}}$  are introduced analogously. Using the shorthand notation  $\Pi_{\mathbf{x}_c} = (\Pi_{n_c}, \Pi_{\mathbf{p}_c}, \Pi_{u_c})$  we also introduced the atomistic expression for the distribution function  $f(\mathbf{x}_c)$  in Eq. (2).

### B. Nonequilibrium Ensemble

Under the assumption that the collection of variables  $(\Pi_{\mathbf{x}_{\text{tot}}}, \Pi_f)$  captures all relevant physical processes on the macroscopic time-scale of interest, the nonequilibrium state of the system is characterized by the generalized canonical ensemble<sup>48,49</sup>:

$$\rho_{\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c} = \frac{1}{Z(\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c)} \delta(\Pi_{\mathbf{x}_{\text{tot}}} - \mathbf{x}_{\text{tot}}) \times e^{-\int d\mathbf{x}_c \boldsymbol{\lambda}_c(\mathbf{x}_c, \mathbf{x}_{\text{tot}}) \Pi_f(\mathbf{x}_c)},\tag{3}$$

where  $Z$  is the partition function of the mixed microcanonical-canonical ensemble  $\rho_{\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c}$ . In particular, the Lagrange multipliers  $\boldsymbol{\lambda}_c$  must be chosen to guarantee for the proper macroscopic averages of the slow variables, i.e.,  $f(\mathbf{x}_c) = \langle \Pi_{f(\mathbf{x}_c)} \rangle_{\rho_{\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c}}$ , where the brackets denote an average over the entire phase space with respect to the ensemble  $\rho_{\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c}$ . We also have  $\mathbf{x}_{\text{tot}} = \langle \Pi_{\mathbf{x}_{\text{tot}}} \rangle_{\rho_{\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c}}$ . While the variables  $\mathbf{x}_{\text{tot}}$  are treated microcanonically since the total system is closed and the quantities  $\mathbf{x}_{\text{tot}}$  rigorously conserved, the nucleus, as an open system, must be treated canonically through the use of Lagrange multipliers.

### C. Entropy

We turn our attention to the entropy of the system described by  $(\mathbf{x}_{\text{tot}}, f(\mathbf{x}_c))$ . Subsequently, the entropy is shown to generate the irreversible dynamics of  $f$ . For the sake of clarity, we have summarized all different entropies involved in our analysis in Table I.

We first derive a useful identity which relies on the observation that the nonequilibrium ensemble can be rewritten as

$$\rho_{\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c} = \frac{1}{Z(\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c)} \delta(\Pi_{\mathbf{x}_{\text{tot}}} - \mathbf{x}_{\text{tot}}) e^{-\boldsymbol{\lambda}_c(\Pi_{\mathbf{x}_c}, \mathbf{x}_{\text{tot}})}, \quad (4)$$

and therefore the distribution  $f$  of the nucleus can be cast into the form

$$\begin{aligned} f(\mathbf{x}_c) &= \langle \Pi_{f(\mathbf{x}_c)} \rangle_{\rho(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)} \\ &= \frac{\tilde{Z}(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}{Z(\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c)} \exp(-\boldsymbol{\lambda}_c(\mathbf{x}_c, \mathbf{x}_{\text{tot}})). \end{aligned} \quad (5)$$

Here,  $\tilde{Z}$  is the microcanonical counterpart of  $Z$ ,

$$\tilde{Z}(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = \int d\bar{\Gamma} \delta(\Pi_{\mathbf{x}_{\text{tot}}} - \mathbf{x}_{\text{tot}}) \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c), \quad (6)$$

and we associate it with the entropy  $S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = k_B \log(\tilde{Z}(\mathbf{x}_{\text{tot}}, \mathbf{x}_c))$  of the system in state  $\mathbf{x}_{\text{tot}}$  with the nucleus of interest having properties  $\mathbf{x}_c$ . We obtain the entropy  $S(\mathbf{x}_{\text{tot}}, f)$  of the system<sup>48,49</sup>

$$\begin{aligned} \frac{S(\mathbf{x}_{\text{tot}}, f)}{k_B} &= \log(Z(\mathbf{x}_{\text{tot}}, \boldsymbol{\lambda}_c)) - \int d\mathbf{x}_c \boldsymbol{\lambda}_c(\mathbf{x}_{\text{tot}}, \mathbf{x}_{\text{tot}}) f(\mathbf{x}_c) \\ &= - \int d\mathbf{x}_c \log\left(\frac{e^{-\boldsymbol{\lambda}_c(\mathbf{x}_{\text{tot}}, \mathbf{x}_{\text{tot}})}}{Z}\right) f(\mathbf{x}_c) \\ &= - \int d\mathbf{x}_c \left[ \log(f(\mathbf{x}_c)) - \frac{S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}{k_B} \right] f(\mathbf{x}_c) \\ &= - \int d\mathbf{x}_c f(\mathbf{x}_c) \log\left(\frac{f(\mathbf{x}_c)}{\phi_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}\right) \\ &\quad + \frac{S_{\text{tot}}(\mathbf{x}_{\text{tot}})}{k_B}, \end{aligned} \quad (7)$$

where  $S_{\text{tot}}(\mathbf{x}_{\text{tot}}) = k_B \log(Z_{\text{tot}})$  is the entropy of the full metastable phase and contains its total partition function

$$\begin{aligned} Z_{\text{tot}} &= \int d\bar{\Gamma} \delta(\Pi_{\mathbf{x}_{\text{tot}}} - \mathbf{x}_{\text{tot}}) \\ &= \int_{\mathbf{x}_{\text{tot}}} d\bar{\Gamma}. \end{aligned} \quad (8)$$

The contribution  $S_{\text{tot}}(\mathbf{x}_{\text{tot}})$  will turn out to be irrelevant in generating the final evolution equations. In the last line of Eq. (8) the integration is over the full phase space compatible with the system state  $\mathbf{x}_{\text{tot}}$ . The expression for the entropy  $S(\mathbf{x}_{\text{tot}}, f)$  contains the entropic potential

$$\begin{aligned} \phi_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) &= \langle \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c) \rangle_{\mathbf{x}_{\text{tot}}} \\ &\equiv \frac{\int_{\mathbf{x}_{\text{tot}}} d\bar{\Gamma} \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c)}{\int_{\mathbf{x}_{\text{tot}}} d\bar{\Gamma}}. \end{aligned} \quad (9)$$

The entropic potential  $\phi_S$  is related to the probability of occurrence of a nucleus with properties  $\mathbf{x}_c$  in a system described by  $\mathbf{x}_{\text{tot}}$ . Upon studying the dynamics of the distribution  $f$  it will turn out that the entropic potential  $\phi_S$  represents the nucleation landscape in which the diffusion process describing nucleation takes place, i.e.,  $V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = -\log(\phi_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c))$ . In the truncated case  $\mathbf{x}_c^{\text{trc}} = (n_c)$  when the full system is isothermal at temperature  $T$  and Brownian motion of the nucleus is irrelevant, the Gibbs-free energy of formation of a nucleus has been recognized<sup>26-32</sup> to be  $k_B T$  times the logarithm of the probability that a nucleus with size  $n_c$  appears in the metastable gas. Our expression  $V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = -\log(\phi_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c))$  hence represents a plausible extension to more nuclei degrees of freedom. In the general setup where the nucleus energy fluctuates, the Gibbs-free energy of formation can no longer be regarded as the proper potential<sup>33</sup>, but now,  $V_S$  or respectively  $S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$  in Eq. (7) as the entropy of formation takes this role. For a collection of the various forms of the nucleation landscape used throughout our work, see Table II.

The entropy of formation  $S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$  is recognized to depend on the nucleus momentum  $\mathbf{p}_c$ . To highlight this feature we show in App. IX A that in the limit of a sufficiently large nucleus ambient phase,

$$\begin{aligned} S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) &= S(\mathbf{x}_{\text{tot}}, n_c, u_c) - \frac{3}{2} k_B n_c \log(n_c) \\ &\quad - m n_c \frac{(\mathbf{v}_c(n_c, \mathbf{p}_c) - \mathbf{v}_{\text{tot}}(N_{\text{tot}}, \mathbf{P}_{\text{tot}}))^2}{2T}, \end{aligned} \quad (10)$$

where  $\mathbf{v}_{\text{tot}} = \mathbf{P}_{\text{tot}}/(mN_{\text{tot}})$  denotes the overall velocity of the system,  $\mathbf{v}_c = \mathbf{p}_c/(m n_c)$  the nucleus velocity and  $S(\mathbf{x}_{\text{tot}}, n_c, u_c)$  represents the entropy of formation if only the variables  $(n_c, u_c)$  characterize the nucleus.

### D. Coarse Grained Evolution Equation

Coarse-grained evolution equations can be split into a reversible part with the system energy as driving force

TABLE I. Summary of relevant entropies.

Expression	Comments
$S(\mathbf{x}_{\text{tot}}, f(\mathbf{x}_c))$	Total system entropy when state $\mathbf{x}_{\text{tot}}$ and nucleus distribution $f(\mathbf{x}_c)$ is known, see Eq. (7). Generates dynamics of $f(\mathbf{x}_c)$ through Eq. (57).
$S_{\text{tot}}(\mathbf{x}_{\text{tot}})$	Entropy of full metastable phase, including the nucleus with arbitrary state.
$S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$	Entropy of system in state $\mathbf{x}_{\text{tot}}$ containing nucleus in state $\mathbf{x}_c$ .
$S(\mathbf{x}_{\text{tot}}, n_c, u_c)$	Entropy of system in state $\mathbf{x}_{\text{tot}}$ containing nucleus in state $(n_c, u_c)$ . Here, the momentum of the nucleus is unknown. See Eq. (10).
$S(\mathbf{x}_{\text{env}}, n_c, u_c)$	Entropy of system, where the system is split into a nucleus in state $(n_c, u_c)$ and its ambient phase in state $\mathbf{x}_{\text{env}}$ . In this set of variables, the entropy can be decomposed: $S(\mathbf{x}_{\text{env}}, n_c, u_c) = S_{\text{env}}(\mathbf{x}_{\text{env}}) + S_c(n_c, u_c)$ .
$S_{\text{env}}(\mathbf{x}_{\text{env}})$	Entropy of nucleus ambient phase to define ambient temperature, pressure and chemical potential, see aSec. IIE.
$S_c(n_c, u_c)$	Entropy of nucleus with state $\mathbf{x}_c$ to define nucleus temperature, see also Sec. (IIF).
$S_{c,K}(n_c, T)$	MFKT entropy of nucleus of size $n_c$ in ambient phase at temperature $T$ .
$S_{c,K}(n_c, u_c)$	Extended MFKT entropy of nucleus in state $(n_c, u_c)$ .

and a part of irreversible origin involving the entropy as driving force<sup>33,49</sup>. For our system the reversible time-evolution is absent since the only involved dynamics is the exchange of mass, momentum and energy of the nuclei with the ambient phase and these processes are known to be purely irreversible in homogeneous media. The fundamental dynamics is hence encompassed in the irreversible contribution to the evolution equations. In App. IX B it is then shown that the evolution equations are recovered as:

$$\begin{aligned} \frac{\partial f}{\partial t} &= \frac{\partial}{\partial \mathbf{x}_c} \cdot \mathbf{D}(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) \cdot \left( \frac{\partial f}{\partial \mathbf{x}_c} + f \frac{\partial V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}{\partial \mathbf{x}_c} \right), \\ \frac{\partial \mathbf{x}_{\text{tot}}}{\partial t} &= 0, \end{aligned} \quad (11)$$

where the nucleation landscape  $V_S$  drives the dynamics and the elements of the diffusion tensor are

$$\mathbf{D}(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = \frac{1}{2k_B \tau_{\text{GK}}} \langle \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \rangle_{(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}, \quad (12)$$

The average  $\langle \dots \rangle_{(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}$  is over all configurations compatible with the state  $\mathbf{x}_{\text{tot}}$  of the total system and the initial state  $\mathbf{x}_c$  of the nucleus. Here,  $\Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} = \Pi_{\mathbf{x}_c}(\tau_{\text{GK}}) - \Pi_{\mathbf{x}_c}(0)$ , where  $\tau_{\text{GK}}$  is a time-scale on which the microscopic correlations drop off and the right hand-side in Eq. (12) reaches a plateau value. For the nucleation of liquid drops in a metastable gas, we recognize  $\tau_{\text{GK}}$  as the intermediate time-scale between the collision time of gas particles with the nucleus and the time-scale on which changes in the nucleus state are recognizable on the coarse-grained level of description. We obtain therefore the tensorial generalization (12) of the expression for the diffusion coefficient  $(2\tau_{\text{GK}})^{-1} \langle \Delta_{\tau_{\text{GK}}} \Pi_{m_c} \Delta_{\tau_{\text{GK}}} \Pi_{m_c} \rangle_{(\mathbf{x}_{\text{tot}}, m_c)}$  in Eq. (1) already used in a variety of papers<sup>50,51</sup> for the truncated state space  $\mathbf{x}_c^{\text{trc}} = (n_c)$ . The evolution equation (11) has been discussed in detail by Schweizer and Sagis<sup>33</sup>, but here we constructed microscopic expressions for its building blocks,  $\mathbf{D}$  and  $V_S$ .

### E. Coarse Grained Evolution Equation in “Standard” Variables

In the literature nucleation is often described by decomposing the system into a nucleating phase and its ambient phase<sup>33,56,57</sup>. We now aim to reformulate our theory in terms of the state variables  $(\mathbf{x}_{\text{env}}, f)$ , where the environment state  $\mathbf{x}_{\text{env}} = (N_{\text{env}}, \mathbf{p}_{\text{env}}, U_{\text{env}}, V_{\text{env}})$  is composed of the total number of particles  $N_{\text{env}}$ , the internal energy  $U_{\text{env}}$ , the momentum  $\mathbf{p}_{\text{env}}$  and the volume  $V_{\text{env}}$  of the nucleus ambient phase. Introducing the average  $\langle A \rangle_f = \int d\mathbf{x}_c f(\mathbf{x}_c) A$  of a coarse-grained variable  $A(\mathbf{x}_c, \mathbf{x}_{\text{tot}})$  with respect to the distribution function  $f$ , we can relate  $\mathbf{x}_{\text{env}}$  to  $\mathbf{x}_{\text{tot}}$ , i.e.,

$$\begin{aligned} V_{\text{env}} &= V_{\text{tot}} - \langle V_c(\mathbf{x}_c) \rangle_f \\ \mathbf{p}_{\text{env}} &= \mathbf{P}_{\text{tot}} - \langle \mathbf{p}_c \rangle_f \\ U_{\text{env}} &= E_{\text{tot}} - \left\langle u_c + \frac{\mathbf{p}_c^2}{2mn_c} \right\rangle_f - \frac{\mathbf{p}_{\text{env}}^2}{2mN_{\text{env}}}, \\ N_{\text{env}} &= N_{\text{tot}} - \langle n_c \rangle_f. \end{aligned} \quad (13)$$

The volume of the ambient phase  $V_{\text{env}}$  is now a time-dependent variable since the volume of the nucleus  $V_c(\mathbf{x}_c)$  continuously changes during the nucleation process. In terms of the new variables the entropy  $S(\mathbf{x}_{\text{env}}, n_c, u_c)$  with eliminated momentum degree of freedom, see also Eq. (10), can naturally be used to define the temperature  $T$  of the ambient phase as the derivative of  $S$  with respect to  $U_{\text{env}}$  at fixed nucleus state, i.e.,  $1/T = \partial S(\mathbf{x}_{\text{env}}, n_c, u_c) / \partial U_{\text{env}}$ . Analogously, the chemical potential  $\mu$  and the pressure

$P$  are given by  $\mu/T = -\partial S(\mathbf{x}_{\text{env}}, n_c, u_c)/\partial N_{\text{env}}$  and  $p/T = \partial S(\mathbf{x}_{\text{env}}, n_c, u_c)/\partial V_{\text{env}}$ . If we now assume that in the variables  $(\mathbf{x}_{\text{env}}, \mathbf{x}_c)$  the entropy splits into its environment and nucleus part, i.e.,  $S(\mathbf{x}_{\text{env}}, n_c, u_c) = S_{\text{env}}(\mathbf{x}_{\text{env}}) + S_c(n_c, u_c)$ , where  $S_{\text{env}}(\mathbf{x}_{\text{env}})$  is a property of the ambient phase only and similarly the nucleus entropy  $S_c(n_c, u_c)$  depends solely on the nucleus state, we find upon transforming the evolution equation (11) to the new variables,

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial \mathbf{x}_c} \cdot \mathbf{D}(\mathbf{x}_{\text{env}}, \mathbf{x}_c) \cdot \left( \frac{\partial f}{\partial \mathbf{x}_c} + f \frac{\partial V_S(\mathbf{x}_{\text{env}}, \mathbf{x}_c)}{\partial \mathbf{x}_c} \right). \quad (14)$$

Nucleation takes now place in the effective nucleation barrier

$$k_B V_S(\mathbf{x}_{\text{env}}, \mathbf{x}_c) = -S_c(n_c, u_c) + V_c \frac{p}{T} - n_c \frac{\mu}{T} + \frac{u_c}{T} + n_c \frac{(\mathbf{v}_c(n_c, \mathbf{p}_c) - \mathbf{v}_{\text{env}}(N_{\text{env}}, \mathbf{p}_{\text{env}}))^2}{2T} + \frac{3}{2} k_B \log(n_c), \quad (15)$$

where  $\mathbf{v}_{\text{env}} = \mathbf{p}_{\text{env}}/(mN_{\text{env}})$  is the velocity of the ambient phase. The expression (15) was already recovered in the nonequilibrium thermodynamics description of the nucleation process and has been discussed in full detail<sup>33</sup>. Our statistical mechanics approach however produces the additional logarithmic contribution in  $n_c$  and the important observation is, that the introduction of the momentum  $\mathbf{p}_c$  into the nucleus state space must be accompanied by this logarithmic term in the nucleation barrier.

## F. Entropic Potential in Thermodynamic Limit

Sampling the entropic potential  $\phi_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$  requires to calculate the probability of occurrence of a nucleus with properties  $\mathbf{x}_c$  in the microcanonical ensemble defined by the state  $\mathbf{x}_{\text{tot}}$ . In the prominent case of a sufficiently large system ensemble averages become identical and instead of averaging over the microcanonical ensemble defined by  $\mathbf{x}_{\text{tot}}$  we can sample over the isobaric-isothermal ensemble with fixed number of particles  $N$ , fixed pressure  $P$  and temperature  $T$ . By Galilean invariance we can set  $\mathbf{P}_{\text{tot}} = 0$  without loss of generality. We define the corresponding system state by  $\tilde{\mathbf{x}}_{\text{tot}} = (N, P, T, \mathbf{P}_{\text{tot}} = 0)$ . Averages of a phase-space variable  $\Pi_A$  with respect to the isobaric-isothermal ensemble are given by

$$\langle \Pi_A \rangle_{\tilde{\mathbf{x}}_{\text{tot}}} = \frac{\int d\bar{\Gamma} \Pi_A e^{-\frac{1}{k_B T} (\Pi_{E_{\text{tot}}} - P \Pi_{V_{\text{tot}}})}}{\int d\bar{\Gamma} e^{-\frac{1}{k_B T} (\Pi_{E_{\text{tot}}} - P \Pi_{V_{\text{tot}}})}}. \quad (16)$$

For the entropic potential (9) we hence obtain the alternative expression

$$\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c) = \langle \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c) \rangle_{\tilde{\mathbf{x}}_{\text{tot}}}, \quad (17)$$

valid when the nucleus ambient phase is sufficiently large. By the same argument other ensembles can be used to

TABLE II. Summary of relevant nucleation landscapes.

Expression	Comments
$V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$	Exact landscape valid for any system size, $-\log \langle \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c) \rangle_{\mathbf{x}_{\text{tot}}}$ .
$V_S(\mathbf{x}_{\text{env}}, \mathbf{x}_c)$	Exact landscape when the system entropy can be split into the nucleus entropy and ambient phase entropy, see Eq. (14). Used to extend MFKT to richer nucleus state space.
$V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)$	Approximate Landscape valid when nucleus ambient phase is large, see Eq. (27). Used to study nucleation when nucleus environment is isothermal and isobaric.
$V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c)$	Approximate truncated landscape valid when nucleus ambient phase is large. Used to study nucleation when nucleus environment is isothermal and isobaric.
$V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^{\text{red}})$	Approximate landscape when nucleus ambient phase is large and the nucleus described by its size and kinetic energy. See Eq. (32). Leads to temperature fluctuation law of McGraw and Laviolette.

sample  $\phi_S$ , such as the canonical or grandcanonical. Only when the system size is small, the rigorous exact version involving microcanonical averaging must be used. The momentum and position degrees of freedom can be separated, so that the entropic potential can be sampled by MC simulations on the position of the particles. To this extent we integrate out the momentum degrees of freedom and rewrite the entropic potential in a more suitable form:

$$\begin{aligned} \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c) &= \int d\mathbf{e}'_c \langle \delta(\Pi_{n_c} - n_c) \delta(\Pi_{\mathbf{p}_c} - \mathbf{p}_c) \\ &\quad \times \delta(\Pi_{u_c^{\text{pot}}} + e'_c - u_c) \\ &\quad \times \delta \left[ \left( \Pi_{e_c^{\text{kin, t}}} - \frac{\mathbf{p}_c^2}{2mn_c} \right) - e'_c \right] \rangle_{\tilde{\mathbf{x}}_{\text{tot}}} \\ &\propto \int_0^\infty d\mathbf{e}'_c F(n_c, e'_c, T) \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c - e'_c), \end{aligned} \quad (18)$$

where

$$F = \frac{1}{n^{3/2} \Gamma(C_V + 1)} \left( \frac{e'_c}{k_B T} \right)^{C_V} e^{-\frac{1}{k_B T} \left( e'_c + \frac{\mathbf{p}_c^2}{2mn_c} \right)} \quad (19)$$

and  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c - e'_c)$  denotes the probability that in the isobaric-isothermal ensemble the nucleus of interest has size  $n_c$  and potential energy  $u_c^{\text{pot}} = u_c - e'_c$ . Here,

$e'_c$  can be thought of an integration over all possible internal kinetic energy contributions. We also introduced the atomistic expression  $\Pi_{u_c^{\text{pot}}}$  of the potential energy of the nucleus as the sum of the potential energies of all atoms of the nucleus, analogous to Eq. (2), and the total kinetic energy  $\Pi_{e_c^{\text{kin, t}}}$  of the nucleus as the sum of the kinetic energy of all its contained atoms. Furthermore,  $C_V = \frac{3}{2}n_c - \frac{5}{2}$  is the heat-capacity of an ideal gas. The probability  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c^{\text{pot}})$  can conveniently be sampled by means of MC simulations since it does not depend on the momentum degrees of freedom.

To gain further insight, we approximate the probability  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c^{\text{pot}})$  by the saddle point approximation obtained for each nucleus size  $n_c$  in the potential minimum  $u_c^{\text{pot},0}(n_c) = \min_{u_c^{\text{pot}}} \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c^{\text{pot}})$ . In this minimum we can evaluate the second derivative of  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c^{\text{pot}})$  with respect to  $u_c^{\text{pot}}$ , which we denote by  $H(n_c)$  and find:

$$\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c^{\text{pot}}) \propto \sqrt{H(n_c)} \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) \times e^{-H(n_c) \frac{(u_c^{\text{pot}} - u_c^{\text{pot},0}(n_c))^2}{2}}. \quad (20)$$

Here the factor of  $\sqrt{H(n_c)}$  is arising for normalization reasons. The entropic potential  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c)$  of the truncated theory is recovered from

$$\begin{aligned} \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) &= \int du_c^{\text{pot}} \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c^{\text{pot}}) \\ &= \int du_c^{\text{pot}} \langle \delta(\Pi_{n_c} - n_c) \delta(\Pi_{u_c^{\text{pot}}} - u_c^{\text{pot}}) \rangle \\ &= \langle \delta(\Pi_{n_c} - n_c) \rangle \end{aligned} \quad (21)$$

and with this we can define the nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) = -\log(\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c))$  of the truncated theory. With the approximation (20) we can solve the integral in Eq. (18) analytically. A more elegant way is to rewrite  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c)$  in the form

$$\begin{aligned} \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c) &\propto \frac{e^{-\frac{p_c^2}{2mn_c k_B T}}}{n_c^{3/2} \Gamma(C_V + 1)} \sqrt{H(n_c)} \phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) \\ &\times \int_0^\infty de'_c e^{-H \frac{(u_c - e'_c - V_0)^2}{2} - \frac{e'_c}{k_B T} + C_V \log\left(\frac{e'_c}{k_B T}\right)}. \end{aligned} \quad (22)$$

We approximate the exponent in the integral to second order around the saddle point  $e_c^s$  and we are then able to calculate the resulting Gaussian integral. The lower integration boundary can be shift to  $-\infty$  since the main contribution of the integral comes from values  $e'_c > 0$  so that the nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c) =$

$-\log(\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c))$  is given by

$$\begin{aligned} V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c) &= V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) + \frac{p_c^2}{2mn_c k_B T} \\ &+ \frac{3}{2} \log(n_c) + \log(\Gamma(C_V + 1)) + H \frac{(u_c - e_c^s - V_0)^2}{2} \\ &+ \frac{e_c^s}{k_B T} - C_V \log\left(\frac{e_c^s}{k_B T}\right) + \frac{1}{2} \log\left(1 + \frac{C_V}{H(e_c^s)^2}\right) \\ &+ \text{const}(T) \end{aligned} \quad (23)$$

that turns out to compare excellently with the true solution (22) of the integral for nucleus sizes down to  $n_c = 5$ , see also Fig 2. To gain further insight, we

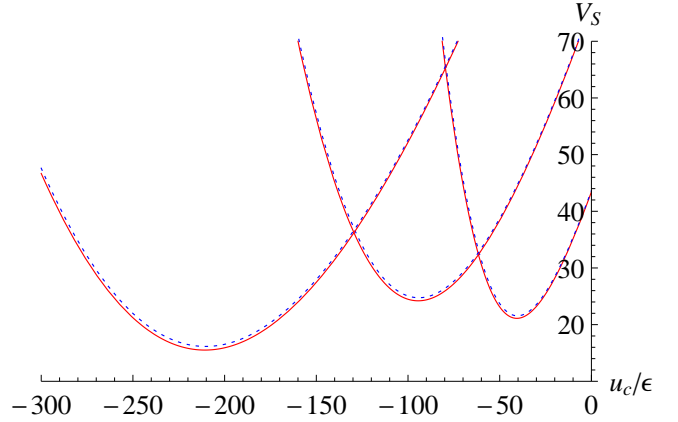


FIG. 2. Nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)$  of a liquid drop in a metastable argon gas for a nucleus with  $n_c = 20, 40$  and  $80$  particles and  $p_c = 0$  as a function of its internal energy  $u_c$ . The red line shows the approximate landscape from Eq. (27) and the dotted blue line the exact landscape with integration of Eq. (22). Thermophysical data is taken from the sixth row in Table V.

can introduce a thermodynamic definition of the nucleus temperature  $T_c$ . We recall that  $V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = -\log(\phi_S(\tilde{\mathbf{x}}_{\text{tot}}(\mathbf{x}_{\text{tot}}), \mathbf{x}_c))$  and upon representing the nucleation landscape  $V_S$  in terms of the state of the ambient phase  $\mathbf{x}_{\text{env}} = (N_{\text{env}}, p_{\text{env}} (= 0), U_{\text{env}}, V_{\text{env}})$ , see Sec. II E, we obtain

$$\begin{aligned} \frac{\partial V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}{\partial \mathbf{x}_c} &= \frac{\partial V_S(\mathbf{x}_e, \mathbf{x}_c)}{\partial \mathbf{x}_c} + \frac{\partial \mathbf{x}_e}{\partial \mathbf{x}_c} \frac{\partial V_S(\mathbf{x}_e, \mathbf{x}_c)}{\partial \mathbf{x}_e} \\ &= \frac{\partial V_S(\mathbf{x}_e, \mathbf{x}_c)}{\partial \mathbf{x}_c} - \frac{\partial V_S(\mathbf{x}_e, \mathbf{x}_c)}{\partial \mathbf{x}_e}. \end{aligned} \quad (24)$$

Notice, that the variable set  $(\mathbf{x}_c, \mathbf{x}_{\text{tot}})$  is more convenient to evaluate the statistical mechanics, but the physical interpretation of  $V_S$  is more appealing in terms of  $(\mathbf{x}_c, \mathbf{x}_e)$ . When changing the nucleation landscape  $V_S(\mathbf{x}_e, \mathbf{x}_c)$  with respect to the internal energy  $U_{\text{env}}$  of the ambient phase, while fixing the nucleus properties  $\mathbf{x}_c$ , this change must be equal to the temperature of the environment, such that  $\partial V_S(\mathbf{x}_e, \mathbf{x}_c) / \partial U_e = -1/k_B T$ , where the minus sign



is a result of our sign convention entering the definition of  $V_S$  through  $V_S = -\log(\phi_S)$ , see also the form of  $V_S$  presented in Eq. (15). Similarly, the nucleus temperature can be assigned as  $\partial V_S(\mathbf{x}_e, \mathbf{x}_c)/\partial u_c = -1/k_B T_c$ . The change of variables law (24) implies

$$\frac{\partial V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}{\partial u_c} = \frac{1}{k_B T} - \frac{1}{k_B T_c} \quad (25)$$

and employing now the specific expression (23) for the purpose of evaluating the left hand-side of Eq. (25) we find the caloric equation of state,

$$u_c = C_V k_B T_c + V_c^0 + \frac{1}{k_B H} \left( \frac{1}{T} - \frac{1}{T_c} \right), \quad (26)$$

and with this, the saddle point can be represented as  $e_c^s = C_V T_c$ . To obtain Eq. (26) we ignored the second last term in Eq. (23) since compared to the other terms it is of subleading order in  $n_c$ . The internal energy of a nucleus is hence given by the kinetic contribution  $C_V T_c$  and a potential energy contribution involving the most probable potential energy  $V_c^0$  at a given nucleus size  $n_c$ , and a correction for nonisothermal effects. When  $T = T_c$ , the nucleus is in the potential minimum of the landscape  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)$  at fixed  $n_c$  and  $\mathbf{p}_c$ . Plugging Eq. (26) and  $e_c^s = C_V T_c$  back into the expression (23) for the entropic potential and expanding the logarithm of the  $\Gamma$ -function with Stirling's approximation, we obtain the final form for the nucleation landscape

$$\begin{aligned} V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c) = & V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) + \frac{\mathbf{p}_c^2}{2Tn_c} \\ & - C_v \left[ \log\left(\frac{T_c}{T}\right) + \left(1 - \frac{T_c}{T}\right) \right] \\ & + \frac{1}{2H} \left( \frac{1}{T} - \frac{1}{T_c} \right)^2 + 2\log(n_c) \\ & + \mathcal{O}\left(\frac{1}{n_c}\right) + \text{const}(T). \end{aligned} \quad (27)$$

Note,  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c)$  is the usual barrier to nucleation in the absence of Brownian-motion and non-isothermal effects. In particular,  $3/2\log(n_c)$  of the correction are due to the introduction of the momentum  $\mathbf{p}_c$  into the nucleus state space and  $1/2\log(n_c)$  through the Stirling approximation on the Gamma function  $\Gamma(C_V + 1)$ . We here recover all corrections reported in the literature, including the momentum contribution obtained through entirely thermodynamic arguments by Reguera and Rubi<sup>44</sup>, and the temperature fluctuation law noticed by McGraw and Laviolette<sup>52</sup>. The additional temperature fluctuation contribution proportional to  $1/H$  has not been reported in the literature to our knowledge. In the subsequent studied model system of argon nucleation the contribution proportional to  $1/H$  effects the energetic landscape of  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)$  dramatically and is hence relevant (see Fig. 2). We postpone further discussion of the temperature fluctuations to Sec. IV.

### III. KINETIC ENERGY AS A STATE-VARIABLE

In view of the various equivalent expressions for the nucleation landscape  $V_S$ , see Eq. (15) and Eq. (27) for instance, it is plausible to ask whether the kinetic energy of the nucleus  $e_c^{\text{kin},s}$  can be used as a state variable instead of the nucleus momentum  $\mathbf{p}_c$ , since the latter enters the entropic barrier only in the contribution  $\mathbf{p}_c^2/2mn_c$  being the “systematic” kinetic energy of the nucleus. Suppose for the further considerations that  $\mathbf{P}_{\text{tot}} = 0$  by Galilean invariance. If we use the modified set of nucleus state variables  $\mathbf{x}_c^{\text{mod}} = (n_c, e_c^{\text{kin},s}, u_c)$  instead of  $\mathbf{x}_c = (n_c, \mathbf{p}_c, u_c)$  we have to construct the entropic potential in terms of these new variables. The entropic potential  $\phi(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$  is the probability that we encounter a nucleus with state  $\mathbf{x}_c$  and therefore we find the clarifying relationship

$$\phi(\mathbf{x}_{\text{tot}}, \mathbf{x}_c^{\text{mod}}) = \int d^3\mathbf{p}_c \delta\left(e_c^{\text{kin},s} - \frac{\mathbf{p}_c^2}{2mn_c}\right) \phi(\mathbf{x}_{\text{tot}}, \mathbf{x}_c). \quad (28)$$

This can be used to obtain the landscape  $V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c^{\text{mod}}) = -\log(\phi(\mathbf{x}_{\text{tot}}, \mathbf{x}_c^{\text{mod}}))$  which looks exactly like  $V_S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)$ , but with  $\mathbf{p}_c^2/2mn_c$  replaced by the new independent state variable  $e_c^{\text{kin},s}$ . The modified set of state variables have the advantage that the state space is smaller upon catching the same relevant macro-physics. The intuition is that in a homogeneous medium it must be irrelevant in which direction the momentum  $\mathbf{p}_c$  of the nucleus points and therefore the nucleation process can equally be described by  $|\mathbf{p}_c|$ ,  $\mathbf{p}_c^2$  or the kinetic energy  $e_c^{\text{kin},s}$ .

### IV. KINETIC ENERGY FLUCTUATIONS IN LITERATURE

It has been convincingly demonstrated in MD simulations<sup>53,54</sup> of nucleation that when only the properties  $\mathbf{x}_c^{\text{red}} = (n_c, e_c^{\text{kin},t})$  of a nucleus are measured, where now  $e_c^{\text{kin},t}$  is the total kinetic energy composed of internal and systematic contributions, the nucleus temperature  $T_c$  can be defined by  $T_c = e_c^{\text{kin},t}/C_V$  and its fluctuations obey the law derived by McGraw and Laviolette,  $\exp\{C_v [\log(T_c/T) + (1 - T_c/T)]\}$ , without the additional contribution involving  $H$  we obtained in Eq. (27). We first remark, that the state space  $\mathbf{x}_c^{\text{red}}$  is a reduced version of  $\mathbf{x}_c$  and therefore cannot capture entirely the rich physics contained in  $\mathbf{x}_c$ . In  $\mathbf{x}_c$  we have additional knowledge of the potential energy contributions to the nucleus internal energy and it is then natural, that these contributions give rise to the more complicated caloric equation of state (26) rather than  $e_c^{\text{kin},t} = C_V T_c$ . Thus, in the reduced framework  $\mathbf{x}_c^{\text{red}}$  we measure kinetic energy fluctuations only, and define the temperature therein while in the richer theory  $\mathbf{x}_c$  we measure the internal energy fluctuations composed of kinetic and potential energy fluctuations, and so it is not surprising that the fluctuation law contains additional contributions.

We apply a similar reduction of degrees of freedom as in Sec. III to show that in the reduced state space  $\mathbf{x}_c^{\text{red}}$  where only the size and total kinetic energy is measured in a simulation, the temperature indeed obeys the equation of state  $e_c^{\text{kin}, t} = C_V T_c$  with the fluctuation law by McGraw and Laviolette. To this extent, notice that the integrand in Eq. (18), denoted by  $\phi(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c, e_c^{\text{kin}, t})$  represents the probability that we observe a nucleus with property  $\mathbf{x}_c$  and total kinetic energy  $e_c^{\text{kin}, t}$ . Therefore

$$\phi(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^{\text{red}}) = \int d^3 \mathbf{p}_c du_c \phi(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c, e_c^{\text{kin}, t}). \quad (29)$$

Employing the approximation (20) the probability  $\phi(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c, e_c^{\text{kin}, t})$  has the form of the right hand side of Eq. (22), but without the integration over  $e_c'$ . Therefore the integrations in Eq. (29) can be performed:

$$\phi(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^{\text{red}}) \propto \frac{\phi(\tilde{\mathbf{x}}_{\text{tot}}, n_c)}{\Gamma(C_V + 1)} e^{-\frac{e_c^{\text{kin}, t}}{k_B T} + C_V \log\left(\frac{e_c^{\text{kin}, t}}{k_B T}\right)}. \quad (30)$$

so that the nucleation landscape  $V_S$  of the reduced theory satisfies up to irrelevant constants

$$V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^{\text{red}}) = V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) + \log(\Gamma(C_V + 1)) - \frac{e_c^{\text{kin}, t}}{k_B T} + C_V \log\left(\frac{e_c^{\text{kin}, t}}{k_B T}\right). \quad (31)$$

The temperature can be defined analogously to the procedure in Sec. IIF and we obtain  $e_c^{\text{kin}, t} = C_V T_c$  and by employing the Stirling approximation finally the suggested fluctuation law by McGraw and Laviolette,

$$V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^{\text{red}}) = V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) + \frac{1}{2} \log(n_c) - C_v \left[ \log\left(\frac{T_c}{T}\right) + \left(1 - \frac{T_c}{T}\right) \right]. \quad (32)$$

This shows that the kinetic energy of the nucleus is indeed a valid measure to obtain the nucleus temperature if no information about its potential energy is available.

## V. APPLICATION: EXTENSION OF MFKT

At the example of the MFKT of nucleation<sup>13</sup> we illustrate that our work represents a powerful tool to extend existing nucleation theories to richer state spaces.

In his work<sup>25</sup>, Kalikmanov derived a non-perturbative expression for the partition function  $Z_{c,K}(n_c, T)$  of a nucleus by microscopic considerations. In his treatment the ambient phase acts as a mean-field on the nucleus. His theory is valid for isothermal nucleation at temperature  $T$  and in particular he finds

$$\begin{aligned} \log(Z_{c,K}(n_c, T)) = & -n_c \frac{\mu_{\text{sat}}(T)}{T} \\ & -k_B [n_c^s(n_c, T) - 1] \theta_{\text{micro}}(T) \\ & + k_B \log(V_{\text{tot}} \rho_{\text{sat}}(T)), \end{aligned} \quad (33)$$

where the chemical potential  $\mu_{\text{sat}}(T)$  and particle density  $\rho_{\text{sat}}(T)$  at saturation are introduced. Here,  $n_c^s$  can be interpreted as the number of particles located at the surface of the nucleus and  $\theta_{\text{micro}}$  is a “microscopic” surface tension.

The entropy  $S_{c,K}(n_c, T)$  and internal energy  $u_c(n_c, T)$  of the nucleus, see also Sec. IIE, are then given by

$$\begin{aligned} S_{c,K}(n_c, T) &= k_B \log(Z_{c,K}(n_c, T)) + k_B T \frac{\partial Z_{c,K}(n_c, T)}{\partial T}, \\ u_c(n_c, T) &= k_B T^2 \frac{\partial Z_{c,K}(n_c, T)}{\partial T}, \end{aligned} \quad (34)$$

by means of the usual definition of statistical mechanics.

### A. Enhancing State Space of MFKT

We exploit a local equilibrium hypothesis to extend the entropy obtained by Kalikmanov to nonisothermal nucleation. To this extent, notice that the entropy  $S_{c,K}(n_c, T)$  of a nucleus can be viewed as a property of the nucleus only and in this sense does not depend on the state of the environment. In isothermal nucleation the temperature of the nucleus  $T_c$  must coincide with the ambient temperature, so that  $T = T_c$  and we can equally write  $S_{c,K} = S_{c,K}(n_c, T_c)$ . Now the local equilibrium hypothesis enters and we postulate that a nucleus in nonisothermal nucleation that can be assigned the temperature  $T_c$  has the same properties as a nucleus in isothermal nucleation that takes place at temperature  $T = T_c$ . Since the entropy, and analogously the internal energy of the nucleus are solely a property of the nucleus and do only depend on the nucleus state, we hence find the generalization of the nucleus entropy to describe nonisothermal nucleation,

$$\begin{aligned} S_{c,K}(n_c, T_c) &= k_B \log(Z_{c,K}(n_c, T_c)) + k_B T_c \frac{\partial Z_{c,K}(n_c, T_c)}{\partial T_c}, \\ u_c(n_c, T_c) &= k_B T_c^2 \frac{\partial Z_{c,K}(n_c, T_c)}{\partial T_c}, \end{aligned} \quad (35)$$

where the nucleus temperature  $T_c$  is now related to its internal energy and mass by  $1/T_c(n_c, u_c) = \partial S_K^{\text{ext}} / \partial u_c$ . This then implies  $S_{c,K}^{\text{ext}}(n_c, u_c) = S_{c,K}^{\text{ext}}(n_c, T_c(n_c, u_c))$ .

### B. Analytic Solution for Nucleation Rate

The entropy  $S_K^{\text{ext}}(n_c, u_c)$  does not depend on the momentum variable  $\mathbf{p}_c$ , so that we use the equations stated in Sec. IIE to obtain the dynamics of extended MFKT. Our description then also takes naturally Brownian motion into account. We take the diffusion tensor of the

particular form<sup>33</sup>

$$\mathbf{D}(n_c) = \begin{pmatrix} \xi(n_c)\mathbf{I}_{3 \times 3} & 0 \\ 0 & \tilde{\mathbf{D}} \end{pmatrix}, \quad (36)$$

and  $\tilde{\mathbf{D}}$  is a  $2 \times 2$  matrix with elements  $\tilde{D}_{n_c, n_c}$ ,  $\tilde{D}_{n_c, u_c}$  and  $\tilde{D}_{u_c, u_c}$  regulating the mass and heat exchange mechanism between nucleus and ambient phase. The evolution equations (14) of a nucleus in a metastable phase at temperature  $T$  and pressure  $P$  can be solved in the saddle point approximation<sup>33,36,58</sup> and give the prediction  $J_{K, \text{ext}}$  of our extended MFKT:

$$J_{K, \text{ext}} = \sqrt{-\frac{1}{2\pi} \frac{\partial^2 V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)}{\partial n_c^2} \bigg|_{\mathbf{x}_c^*}} \rho_{g, \tilde{\mathbf{x}}_{\text{tot}}} \times \frac{1}{\omega} \frac{|\lambda_{\min}|}{\sqrt{|\det \mathbf{G}|}} e^{-\Delta V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^*)}, \quad (37)$$

where

$$\mathbf{G} = \frac{\tilde{\zeta}^2}{\sigma_{u_c}^2} \begin{pmatrix} 1 - \frac{\omega^2}{\tilde{\zeta}^2} & \frac{1}{\tilde{\zeta}} \\ \frac{1}{\tilde{\zeta}} & \frac{1}{\tilde{\zeta}^2} \end{pmatrix}. \quad (38)$$

The coefficients are given by

$$\begin{aligned} \sigma_{u_c}^2 &= T^2 \frac{\partial u_c(n_c, T)}{\partial T} \bigg|_{n_c^*}, \\ \frac{1}{\sigma_{n_c}^2} &= -\theta_{\text{micro}}(T) \frac{\partial^2 n_s(T, n_c)}{\partial n_c^2} \bigg|_{n_c^*}, \\ \omega^2 &= \frac{\sigma_{u_c}^2}{\sigma_{n_c}^2} \bigg|_{n_c^*}, \\ \tilde{\zeta} &= -c_v T + T^2 \frac{\partial \log(\rho_{\text{sat}}(T))}{\partial T} + T^2 \frac{\partial^2 n_s(T, n)}{\partial n_c \partial T} \bigg|_{n_c^*} \theta_{\text{micro}} \\ &\quad + T^2 \left| \frac{\partial n_s(T, n)}{\partial n_c} \right|_{n_c^*} \frac{d\theta_{\text{micro}}(T)}{dT}, \end{aligned} \quad (39)$$

and

$$\lambda_{\min} = \frac{1}{2} \left( \text{tr} \mathbf{R} - \sqrt{(\text{tr} \mathbf{R})^2 - 4 \det \mathbf{R}} \right). \quad (40)$$

identified as the smallest eigenvalue of  $\mathbf{R} = \tilde{\mathbf{D}}\mathbf{G}$ . All quantities in Eq. (39) and  $\lambda_{\min}$  must be evaluated at the saddle point  $\mathbf{x}_c^* = (n_c^*, \mathbf{p}_c^* = 0, u_c^*)$  of the nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)$  as given by Eq. (15). For the internal energy  $u_c(n_c, T_c)$  this saddle point falls on  $u_c^* = u_c(n_c^*, T)$ . Here,  $\Delta V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^*)$  denotes the height of the nucleation landscape and  $\rho_g$  is the particle density of the metastable phase in which nuclei are formed. In the saddle point approximation (37) the effects of Brownian motion are seemingly irrelevant since the coefficient

$\xi$  is not involved. This is however not true since the logarithmic normalization factor  $3/2 \log(n_c^*)$  contained in  $\Delta V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c^*)$  enters the prediction.

For the prediction of the standard isothermal MFKT nucleation rate  $J_K$  we refer to the book of Kalikmanov<sup>13</sup> (Chapter 7.7 Steady State Nucleation Rate).

### C. Application to Nucleation of Argon

As a specific model system we consider the nucleation of argon droplets in a supersaturated gas in the low-supersaturation regime studied by Diemand et al. through brute-force MD simulations<sup>6</sup>. This allows to compare our predictions directly with first hand exact results on the nucleation rate. For the physical setups listed in Table IV we have computed the isothermal MFKT nucleation rate  $J_K$  and the corrected rate  $J_{K, \text{ext}}$  and compared it to the MD simulations by Diemand. Since we compared our nucleation rate predictions to MD simulations, we did not take the real thermophysical data of argon, but data obtained through MD simulations<sup>6</sup>, see also Table III.

In Fig. 5 we illustrated the predictions of the standard isothermal MFKT and the extended version. The truncated theory seems to deviate systematically from the exact MD rates  $J_{\text{MD}}$ . Our theory successfully eliminates this deviation indicating that we chose the proper set of variables to capture the nucleation process on the macroscale. Notice, that we did not include effects of angular momentum and shape fluctuations in our description. As argued by Angéilil et al.<sup>54</sup> the effect of angular momentum seems only to be pronounced for small nuclei consisting of less than 16 atoms and all our critical sizes were beyond this size, so that angular momentum should not be used as an independent state variable for an autonomous description of the nucleation process in the low supersaturation regime. Angéilil et al. recognized that the shape of large nuclei can deviate significantly from spherical objects. Using the shape of a nucleus as an independent fluctuating variable might offer additional insights into the relevance of this correction.

Specifically, we evaluated the MFKT nucleation rate by resampling to the following relationships for which precise MD thermophysical data was available, see Table III. We shortly summarize the parameters appearing in the MFKT and closely follow the lines of its inventor<sup>13,25</sup>. According to Kalikmanov, the microscopic surface tension is given by  $\theta_{\text{micro}}(T_c) = -\log(-B_2(T_c)p_{\text{sat}}(T_c)/(k_B T_c))$ , where  $B_2$  is the second virial coefficient of the gas. Introducing the macroscopic surface tension  $\gamma_S(T_c)$ , we can define the ratio  $\tilde{\omega}(T_c) = \gamma_S(T_c)/(k_B T_c \theta_{\text{micro}}(T_c))$ . We further define  $\tilde{\zeta} = n_c^{-1/3}$  and the coefficient  $\tilde{\lambda} = \sqrt{N_1/\tilde{\omega} - 3/4} - 3/2$ , where  $N_1$  is the number of neighbours of particles in the nucleus. The number of neighbours in the nucleus is evaluated from the argon-liquid properties and can be approximated by  $N_1(T_c) = 5.5116\eta(T_c)^2 + 6.1383\eta(T_c) + 1.275$ , where

$\eta(T_c) = \frac{\pi}{6} \rho_l(T_c) d_{\text{hs}}^3$ . Here,  $d_{\text{hs}}$  is given by the Barker-Henderson relation with the Weeks-Chandler-Anderson decomposition of the intermolecular potential, and represents the effective hard sphere diameter in the theory of liquids,  $d_{\text{hs}}(T_c) = a_1 T + b_1 / (a_2 T_c + a_3) \sigma$  with the Lennard-Jones parameters  $\sigma = 3.822 \times 10^{-10}$  m and  $\epsilon = 1.654 \times 10^{-21}$  J. The coefficients are given by  $a_1 = 0.56165 \frac{k_B}{\epsilon}$ ,  $a_2 = 0.60899 \frac{k_B}{\epsilon}$ ,  $a_3 = 0.92868 \frac{k_B}{\epsilon}$  and  $b_1 = 0.9718 \frac{k_B}{\epsilon}$ . The mass of argon particles is  $m = 6.690 \times 10^{-26}$  kg. The average number of particles  $n_c^s(n_c, T_c) = \tilde{\alpha}(n_c, T_c) n_c$  located at the surface of the nucleus is obtained by solving a cubic equation for the introduced coefficient  $\alpha(n_c, T_c)$ . In particular,  $\tilde{\alpha} = 3\omega\tilde{\zeta}(1-\alpha)^{2/3} + 3\tilde{\omega}\tilde{\lambda}\tilde{\zeta}^2(1-\alpha)^{1/3} + \tilde{\zeta}^3\tilde{\omega}\tilde{\lambda}^2$ . For the

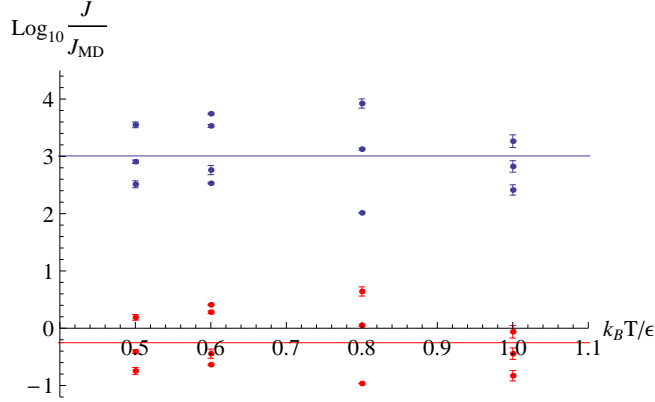


FIG. 3. Comparison of argon nucleation rates  $J$  obtained from recent large-scale molecular dynamics simulations<sup>6</sup>,  $J_{\text{MD}}$ , with the original mean-field kinetic theory<sup>25</sup> (blue) and our extension (red). At each temperature  $T$  the rate comparison for various supersaturation ratios  $S$  is presented, see also Table IV. The solid lines are a guide to the eyes.

kinetic coefficient  $\tilde{D}_{n_c, n_c}$  we used the standard formula obtained from kinetic theory<sup>35,36</sup>

$$\tilde{D}_{n_c, n_c} = \frac{p_{\text{sat}}}{\sqrt{2\pi m k_B T}} A(n_c), \quad (41)$$

where  $A(n_c) = 4\pi r_c^2$  is the surface area of the nucleus and  $r_c = (3\rho_l n_c / 4\pi)^{1/3}$  its radius. The remaining coefficients are given by predictions of Barret<sup>33,36</sup>, i.e.,  $\tilde{D}_{n_c, u_c} = 2k_B T \tilde{D}_{n_c, n_c}$  and  $\tilde{D}_{u_c, u_c} = 6(k_B T)^2 \tilde{D}_{n_c, n_c}$  and are a result of kinetic considerations of the microscopic mass and energy exchange process between ambient phase and nucleus.

## VI. APPLICATION: NUCLEATION IN LENNARD-JONES GAS

We study the nucleation of liquid drops in a metastable Lennard-Jones gas in the low supersaturation regime considered by the latest large-scale MD simulations<sup>6</sup> by Diemand et al. In order to compare our nucleation rate predictions obtained from our coarse-graining procedure

TABLE III. Thermophysical data of argon<sup>6</sup>.

$k_B T / \epsilon$	$p_{\text{sat}} \sigma^3 / \epsilon$	$\gamma_S \sigma^2 / \epsilon$	$\rho_l \sigma^3 / m$	$B_2 / \sigma^3$
1.0	$2.55 \times 10^{-2}$	0.453	0.696	-5.26
0.8	$4.53 \times 10^{-3}$	0.863	0.797	-7.75
0.6	$2.54 \times 10^{-4}$	0.882	0.882	-12.9
0.5	$2.54 \times 10^{-5}$	0.921	0.921	-18.15

TABLE IV. Nucleation rate predictions of argon at various temperature  $T$  and supersaturations  $S$ . The initial density of the metastable gas is  $\rho_g$ . Shown are the predictions  $J_{\text{MD}}$  of the brute-force MD simulations<sup>6</sup> by Diemand et al. the predictions of the original MFKT<sup>25</sup>,  $J_K$ , and the predictions,  $J_{K, \text{ext}}$  of our extension of MFKT.

$k_B T / \epsilon$	$S$	$\rho_g \sigma^3$	$\log_{10} \left( \frac{J_{\text{MD}}}{\sigma^3 \tau} \right)$	$\log_{10} \left( \frac{J_K}{\sigma^3 \tau} \right)$	$\log_{10} \left( \frac{J_{K, \text{ext}}}{\sigma^3 \tau} \right)$
0.5	72.8	0.008	$-11.36 \pm 0.03$	-9.15	-12.14
0.5	59.2	0.007	$-13.01 \pm 0.02$	-10.41	-13.44
0.5	48.7	0.005	$-15.07 \pm 0.04$	-11.83	-14.92
0.6	16.9	0.008	$-11.75 \pm 0.01$	-9.53	-12.42
0.6	15.6	0.007	$-12.64 \pm 0.02$	-10.17	-13.10
0.6	14.0	0.006	$-14.39 \pm 0.05$	-11.16	-14.15
0.6	12.0	0.005	$-16.32 \pm 0.73$	-12.88	-15.93
0.8	4.02	0.030	$-9.08 \pm 0.01$	-7.36	-10.08
0.8	3.55	0.025	$-11.70 \pm 0.01$	-8.87	-11.70
1.0	1.66	0.062	$-10.94 \pm 0.01$	-8.83	-11.79
1.0	1.63	0.060	$-11.90 \pm 0.07$	-9.38	-12.36
1.0	1.60	0.058	$-12.92 \pm 0.07$	-9.96	-13.00

with the MD simulations, we employ the same Lennard-Jones parameters as Diemand et al. and consider here nucleation in the limit of a large ambient phase at temperature  $T$  and pressure  $P$ . We aim to obtain the static building block  $\phi_S(\mathbf{x}_c, \tilde{\mathbf{x}}_{\text{tot}})$  and the dynamic information  $\mathbf{D}_{\mathbf{x}_c, \mathbf{x}_c}$  from atomistic simulations. This then determines the coarse-grained model. We apply the following scheme to obtain predictions of the nucleation rate:

1. Sample configurations  $C$  distributed according to the isothermal-isobaric ensemble in Eq. (16) through MC simulations to obtain  $\phi_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c, u_c^{\text{pot}})$ , here approximated by Eq. (20). Construct the nucleation barrier  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)$  in Eq. (27).
2. Use the configurations  $C$  as input in short-time MD simulations to obtain  $\mathbf{D}_{\mathbf{x}_c, \mathbf{x}_c}$ .
3. Solve the Fokker-Planck equation (11) in the stationary limit to obtain the nucleation rate predic-

tion.

Subsequently we offer more details to each of the steps.

### A. Lennard-Jones Model System

We simulate the interaction between atoms through the long-ranged Lennard-Jones potential, expressed at a separating distance  $r$  by

$$\phi(r) = \begin{cases} 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] & \text{if } 0 < r < r_c, \\ 0 & \text{otherwise.} \end{cases}$$

The cutoff radius is  $r_c = 5 \times \sigma$ . Typical values of  $\sigma$  and  $\epsilon$  for a monoatomic Lennard-Jones system, made of argon-like particles, are  $\sigma = 3.822 \times 10^{-10}$  m and  $\epsilon = 1.654 \times 10^{-21}$  J, with a particle's mass  $m = 6.690 \times 10^{-26}$  kg. We also introduce the typical time-scale  $\tau = \sqrt{\sigma^2 m / \epsilon}$  on which the atomistic dynamics takes place. We performed both, the MC and MD simulations with  $N_{\text{tot}} = 1000$  particles in the isobaric-isothermal ensemble. We used periodic boundary conditions in all directions. With 1000 particles system size effects are negligible and we benefit from the fact that the system is small enough so that only one nucleus is formed in the system.

### B. Clustering of Particles

For our simulations we need to formulate the microscopic cluster criterion  $\Pi_{\theta(\bar{\Gamma})}$  introduced in Sec. II A to evaluate nucleus properties. We here rely on the extension<sup>32</sup> of ten Wolde et al. of the most popular and simple scheme, the Stillinger-cluster definition where nearby atoms, within a distance smaller than a cutoff distance  $r_c = 1.5\sigma$  are considered part of the same cluster. Ten Wolde et al. further distinguish between particles with a liquid like environment and vapor like environment. A particle has a liquid like environment if it has at least 5 neighbours as defined by the Stillinger-criterion. All liquid like particles which are connected are considered part of the same nucleus. The scheme by ten Wolde et al. is computationally cheap and has been shown by Wedekind et al.<sup>55</sup> to offer several advantages over the simple Stillinger definition. The developed nonequilibrium statistical mechanics does not indicate which cluster criterion should be used, but once the cluster criterion is fixed, the formulation of our coarse-graining procedure is independent of it.

### C. MC Simulations

To sample the entropic potential  $\phi_S$  we exploit standard MC simulations<sup>55</sup> in the isobaric-isothermal ensemble with fixed number of particles  $N$ , pressure  $P$  and temperature  $T$ . The formation of a nucleus in the low

supersaturation regime of the Lennard-Jones gas is a rare-event and so we resample to an importance sampling technique to sample regions in phase space that contain a nucleus. We use the umbrella-sampling technique to bias the simulation with the bias potential

$$\Pi_{n_c, \text{bias}} = -\frac{1}{2}k(\Pi_{n_c} - n_c)^2 \quad (42)$$

that favours the presence of a nucleus with size  $n_c$  in the system. We use different umbrella windows corresponding to different choices of  $n_c$ . In particular, we take umbrella windows from cluster size 10 to 140 usually in steps of 10. The bias strength is chosen as  $k = 0.25$ . The cluster size 140 is in all simulations beyond the critical size  $n_c^*$ . In our MC simulations of the  $NPT$  ensemble, we attempt at each step a volume changing move with probability  $1/N$ . If the volume changing move does not take place, each particle is moved with a probability of  $5/N$  in this step. The maximum distance in any direction a particle can move is set to  $0.4\sigma$ . This ensures in all our simulations an average acceptance probability of volume moves and particle moves of 50%. To sample each umbrella window we first initialize homogeneously distributed particles in space so that their density coincides with the density of the metastable state at the temperature and supersaturation we are investigating, see also Table V. Subsequently we perform an equilibration run of  $10^7$  MC steps and then start averaging for  $10^8$  MC steps to obtain the entropic potential. The cluster crite-

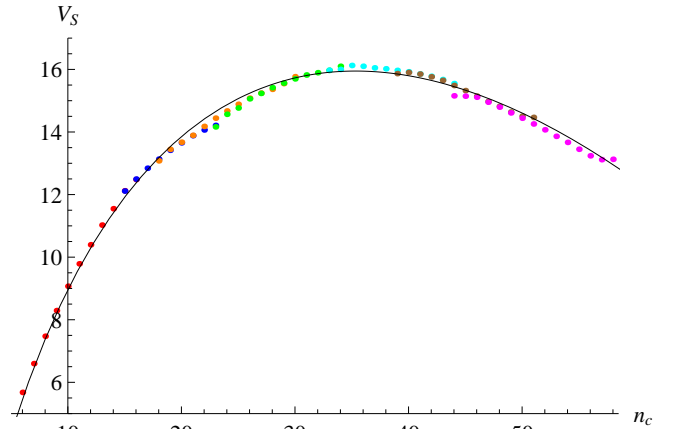


FIG. 4. Nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c)$  for argon at temperature  $T = 0.7\epsilon/k_B$  and supersaturation  $S = 6.11$  near the critical size  $n_c \approx 35$ . Dots are obtained from Monte-Carlo simulations and the different colors indicate different umbrella windows. The solid line is a fit with Eq. (43).

riion of ten Wolde et al. allows for a nucleus larger than 6 particles only. We however fit the data obtained for  $\phi_S$  between  $n_c = 6$  and 140 and assume this fit can be extrapolated down to  $n_c = 1$ . We found excellent fits of

the nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, \mathbf{x}_c)$  by Eq. (27) with

$$V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c) = -\log(S)(n_c - 1) + \eta_1 \left( n_c^{\frac{2}{3}} - 1 \right) + \eta_2 \left( n_c^{\frac{1}{3}} - 1 \right), \quad (43)$$

where  $\eta_1$  and  $\eta_2$  depend on the temperature  $T$  and supersaturation  $S$ . Furthermore, we fit the potential minimum  $u_c^{\text{pot},0}(n_c) = e_0 n_c + e_s n_c^{2/3}$  and  $H(n_c) = h_1 n_c^{-3/2}$ , see also Table V. Here,  $e_0$  can be interpreted as the bulk energy per particle of a nucleus, and  $e_s$  as its corresponding surface energy. In Fig. 4 we illustrate a typical nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c)$  near the critical cluster size reconstructed from the MC simulations.

#### D. MD Simulations

The atomic positional configurations harvested from MC simulations are used as initial conditions for our short-time molecular dynamics simulations in the microcanonical ensembles. We can equally use the isobaric-isothermal ensemble at temperature  $T$  and pressure  $P$  to sample the short-time correlations of fluctuations determining the diffusion coefficient since on these short time-scales the macroscopic state of the system cannot be affected. Microcanonical simulations are more robust since there is no thermostat that can affect the nucleus state. If the isobaric-isothermal ensemble is used, the thermostats corresponding to  $T$  and  $P$  must be ensured to produce system relaxation times of the thermostated variables that drop off much slower than the time-scale on which relevant microscopic correlations drop off. Otherwise, the nucleus momentum and energy can be affected through the thermostat on the same time-scale as microscopic collisions of atoms of the ambient phase and the nucleus take place, resulting then in erroneous estimates of the diffusion tensor<sup>55</sup>. For simplicity we have sampled the initial momentum of all atoms in our MD simulations with a Gaussian distribution at the temperature  $T$  we are investigating. In this work we omit studying the energy dependence of the diffusion tensor and solely capture its size dependence. To sample the diffusion tensor at the critical cluster size for any given temperature and supersaturation regime, we take 32 configurations harvested from MC runs that contain a nucleus of the critical size and then run the MD simulation for a total time of  $25\tau$  around five times larger than the collision time-scale  $\tau_{\text{GK}} \approx 5\tau$  of gas particles with the nucleus. By averaging the dynamics of all 32 simulations we are able to obtain the required mean-square displacements (12). The corresponding correlation functions  $\langle \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \rangle_{(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}$  are illustrated in Fig. 5, and within the collision time-scale a linear regime is recognizable from which the diffusion coefficient can be obtained. In Table V we have listed all obtained diffusion coeffi-

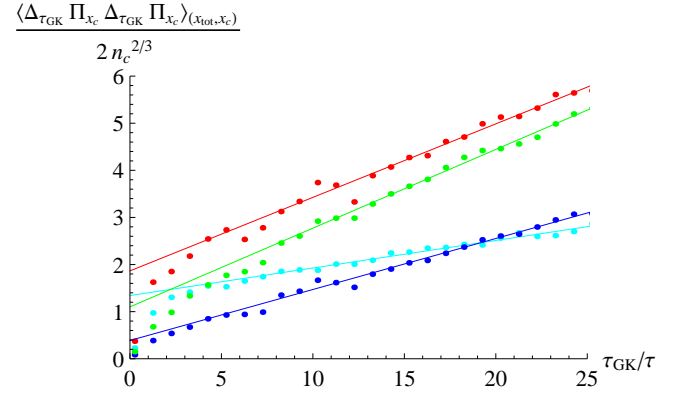


FIG. 5. Diffusion matrix normalized by  $n_c^{2/3}$  for nucleation of argon at temperature  $T$  and supersaturation  $S = 6.11$ . The dots denote the mean-square displacement of  $n_c$  with  $n_c$  (red), between  $n_c$  with  $u_c$  (blue), between  $u_c$  with  $u_c$  and  $\mathbf{p}_c^x$  with  $\mathbf{p}_c^x$ , where  $\mathbf{p}_c^x$  is the momentum in  $x$ -direction. The solid lines, the slope of which enters the prediction of the diffusion tensor, are fits to the linear regime of these correlations.

cients with the convention

$$D(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = \begin{pmatrix} \xi^{r,0} \mathbf{I}_{3 \times 3} & 0 \\ 0 & 1 & D_{n_c, u_c}^{r,0} \\ 0 & D_{n_c, u_c}^{r,0} & D_{u_c, u_c}^{r,0} \end{pmatrix} D^0 n_c^{2/3}. \quad (44)$$

We also evaluated the diffusion coefficients for the temperature  $T = 0.7$  and supersaturation ratio 6.11 for nucleus sizes distinct from the critical size and realized that they are well approximated by Eq. (44) for nuclei down to 15 atoms. The scaling (44) was also valid in the post-critical regime. We hence assume the validity of Eq. (44) for all nucleus sizes. In his work<sup>35,36</sup>, Barret gives explicit expressions for the coefficients  $D_{n_c, u_c}^{r,0}$  and  $D_{u_c, u_c}^{r,0}$  derived by kinetic theory arguments, see also Sec. V C. We did not find his theoretical predictions to match with our computer experiments. But we found rather good agreement of his prediction of  $D^0$  with ours.

In evaluating the coefficients  $\langle \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \rangle_{(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}$  we notice that the microscopic process behind its contribution is the exchange of mass, momentum and energy of the nucleus with the ambient phase. We adopted the cluster criterion of ten Wolde et al. in our MC simulations, and within this criterion there are always particles located at the surface of the nucleus which are not part of the nucleus since they have less than 5 neighbours. These particles can however stay at the surface of the nucleus for a considerable time-scale much larger than the collision time-scale  $\tau_{\text{GK}}$  and even rapidly oscillate between being part of the nucleus and part of the ambient phase. These oscillations lead to significant noise effects that make it difficult to evaluate the true diffusion coefficients, which by nature should only be affected by particles that truly

TABLE V. Fit of the entropic potential (27) and diffusion tensor (44) with conventions and parameters introduced in Sec. VIC and Sec. VID. In  $e_0$  and  $e_s$  as well as  $h_1$  the error is smaller than the number of presented digits. The error in  $D^0\tau$  has been discussed in Sec. VID. Errors in  $D_{n_c, u_c}^{r,0}\epsilon^2$  and  $\xi^{r,0}m\epsilon$  are smaller than 0.05 and errors in  $D_{u_c, u_c}^{r,0}\epsilon^4$  smaller than 0.1.

$k_B T/\epsilon$	$S$	$\rho_g \sigma^2$	$\eta_1$	$\eta_2$	$e_0/\epsilon$	$e_s/\epsilon$	$h_1 \epsilon^2$	$D^0\tau$	$D_{n_c, u_c}^{r,0}\epsilon^2$	$D_{u_c, u_c}^{r,0}\epsilon^4$	$\xi^{r,0}m\epsilon$
0.8	2.98	0.0200	$7.81 \pm 0.04$	$-9.73 \pm 0.25$	-4.8	5.4	6.8	0.089	0.72	1.79	0.77
0.8	3.11	0.0211	$7.86 \pm 0.04$	$-8.99 \pm 0.12$	-4.7	5.3	6.6	0.132	0.71	1.50	0.62
0.8	3.33	0.0230	$7.64 \pm 0.05$	$-7.73 \pm 0.24$	-4.7	5.1	6.6	0.229	0.68	1.15	0.50
0.8	3.55	0.0250	$7.64 \pm 0.07$	$-7.50 \pm 0.31$	-4.6	4.9	6.5	0.314	0.68	0.88	0.39
0.7	5.30	0.0106	$10.62 \pm 0.07$	$-12.19 \pm 0.23$	-4.9	5.1	7.5	0.054	0.75	1.42	0.80
0.7	6.11	0.0125	$10.27 \pm 0.05$	$-9.27 \pm 0.38$	-4.8	4.9	8.5	0.145	0.74	1.16	0.37
0.6	16.90	0.0080	$12.21 \pm 0.06$	$-7.63 \pm 0.70$	-5.3	5.7	4.8	0.056	0.54	0.89	0.49
0.6	15.60	0.0073	$12.04 \pm 0.07$	$-7.26 \pm 0.21$	-5.3	5.8	5.1	0.038	0.61	1.03	0.63
0.6	14.00	0.0065	$11.89 \pm 0.05$	$-6.04 \pm 0.18$	-5.3	5.8	5.3	0.032	0.68	1.20	0.77
0.6	11.95	0.005	$12.06 \pm 0.08$	$-7.19 \pm 0.12$	-5.3	5.7	6.4	0.025	0.74	1.42	0.81
0.5	39.37	0.0021	$16.34 \pm 0.09$	$-11.08 \pm 0.32$	-5.6	6.3	5.1	0.013	0.70	1.00	0.62
0.5	43.31	0.0023	$16.57 \pm 0.10$	$-11.94 \pm 0.28$	-5.6	6.2	5.2	0.012	0.67	0.85	0.56
0.5	48.70	0.0026	$16.74 \pm 0.09$	$-12.53 \pm 0.23$	-5.6	6.3	5.4	0.012	0.58	0.69	0.48

collide, are absorbed or emitted by the nucleus and do not fall in the undesirable class of oscillating surface particles, an inherent drawback of the ten Wolde cluster criterion. The contribution to the diffusion coefficient from the oscillating particles should disappear if we average over sufficiently many simulations. The resulting noise effects are understood by the fact that the number of attempts with which the outer shell particles oscillate between being a nucleus property and a gas property outnumbers the true collision attempts from the ambient phase. We found that we can significantly improve the statistics of  $\langle \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \rangle_{(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}$  by using instead of the ten Wolde cluster criterion, the Stillinger-cluster criterion only so that now the oscillating particles in the ten Wolde cluster definition belong to the nucleus as well and thereby, the nucleus state  $\mathbf{x}_c$  only changes through desirable collision processes. It is expected that this procedure results in an error in the obtained diffusion coefficients since by deciding that the outer layers of the nucleus is now also a nucleus property, the original nucleus state obtained from the ten Wolde definition is modified. This error only affects the prefactor  $D_{n_c, n_c}^0$  in Eq. (44). The Stillinger criterion gives typically nucleus sizes  $n_c^{\text{Stillinger}}$  that are  $3/2$  larger than the ten-Wolde nucleus size  $n_c^{\text{ten Wolde}}$ . In Eq. (44) we set  $D^0 = D_{n_c, n_c}/n_c^{2/3}$ . Suppose we obtained  $D_{n_c, n_c}$  in our simulation around the critical size  $n_c^{*, \text{Stillinger}} \approx 3/2 n_c^{*, \text{Wolde}}$  with the Stillinger criterion, then  $D^0 = D_{n_c, n_c}/(n_c^{\text{Stillinger}})^{2/3}$  and if we use the ten-Wolde definition, we must get the same coefficient  $D_{n_c, n_c}$ , but  $D^0 = D_{n_c, n_c}/(n_c^{\text{Wolde}})^{2/3}$ . This means that the two values obtained for  $D^0$  are distinct by a factor  $(n_c^{*, \text{Stillinger}}/n_c^{*, \text{ten Wolde}})^{3/2} \approx 2$ , so that the nucleation

rate prediction that depends linearly on  $D^0$  is affected by an error of a factor of two. In nucleation theory such errors are completely negligible.

### E. Rate-Estimation

The rate prediction for the truncated theory using  $\mathbf{x}_c^{\text{trc}} = (n_c)$  as state variable is given by the classical formula

$$J_{\text{trc}} = \sqrt{-\frac{1}{2\pi} \frac{\partial^2 V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c)}{\partial n_c^2}} \Big|_{n_c^*} \rho_g \times D^0 (n_c^*)^{\frac{3}{2}} e^{-\Delta V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c^*)}, \quad (45)$$

where  $n_c^*$  is the critical size of the truncated nucleation landscape  $V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c^*)$  and  $\Delta V_S(\tilde{\mathbf{x}}_{\text{tot}}, n_c^*)$  the barrier height, see also Eq. (43). In order to obtain the nucleation rate of the full theory  $J_{\text{ext}}$ , we performed Brownian dynamics (BD) simulations of the full Fokker-Planck equation (11) and the Fokker-Planck equation of the truncated theory (1) to obtain the correction to the nucleation rate (45) through nonisothermal effects and Brownian motion. For the BD simulation we closely follow earlier work<sup>33</sup>. In Table VI we tabulated the different obtained rates and corrections  $\alpha_{\text{non-iso}}$  and  $\alpha_{\text{BM}}$  to  $J_{\text{trc}}$  through nonisothermal and Brownian effects of the nucleus. We roughly estimated these correction by first performing the BD simulations in an isothermal setup, but with the correction for Brownian motion activated to obtain  $\alpha_{\text{BM}}$  and subsequently we performed BD simulations incorporating all state variables and the additional correction was attributed to  $\alpha_{\text{non-iso}}$ . We observed that in



TABLE VI. Nucleation rate predictions for a metastable Lennard-Jones gas at various temperatures  $T$  and supersaturations  $S$ .  $J_{\text{MD}}$  is the prediction of MD simulations<sup>6</sup> (an asterisk on top of the data indicates that the rate is obtained through the universal scaling relation<sup>59</sup>).  $J_{\text{trc}}$  is the prediction of the standard technique using the truncated state space and  $J_{\text{ext}}$  uses the full state space.  $\alpha_{\text{non-iso}}$  and  $\alpha_{\text{BM}}$  estimates the correction of the nucleation rate  $J_{\text{trc}}$  due to nonisothermal effects and Brownian motion of the nucleus respectively.

$k_{\text{B}}T/\epsilon$	$S$	$\log_{10} \left( \frac{J_{\text{MD}}}{\sigma^3 \tau} \right)$	$\log_{10} \left( \frac{J_{\text{trc}}}{\sigma^3 \tau} \right)$	$\log_{10} \left( \frac{J_{\text{ext}}}{\sigma^3 \tau} \right)$	$\log_{10} \left( \frac{\alpha_{\text{non-iso}}}{\sigma^3 \tau} \right)$	$\log_{10} \left( \frac{\alpha_{\text{BM}}}{\sigma^3 \tau} \right)$
0.8	2.98	$-16 \pm 0.5$	$-11.43 \pm 0.29$	$-16.08 \pm 0.61$	-1.0	-3.7
0.8	3.11	$-15.53^*$	$-11.33 \pm 0.20$	$-16.03 \pm 0.42$	-1.0	-3.6
0.8	3.33	$-13.47 \pm 0.07$	$-9.71 \pm 0.27$	$-14.34 \pm 0.58$	-1.2	-3.4
0.8	3.55	$-11.90 \pm 0.02$	$-8.59 \pm 0.33$	$-13.20 \pm 0.60$	-1.4	-3.2
0.7	5.30	$-16.24^*$	$-11.89 \pm 0.31$	$-16.40 \pm 0.59$	-1.2	-3.3
0.7	6.11	$-13.28^*$	$-10.43 \pm 0.30$	$-14.83 \pm 0.68$	-1.3	-3.1
0.6	16.90	$-11.96 \pm 0.01$	$-8.52 \pm 0.30$	$-12.21 \pm 0.62$	-1.4	-2.3
0.6	15.60	$-12.81 \pm 0.04$	$-9.11 \pm 0.16$	$-12.86 \pm 0.35$	-1.3	-2.5
0.6	14.00	$-14.58 \pm 0.07$	$-10.53 \pm 0.14$	$-14.39 \pm 0.29$	-1.3	-2.6
0.6	11.95	$-16.45 \pm 0.98$	$-11.83 \pm 0.18$	$-15.88 \pm 0.36$	-1.2	-2.8
0.5	39.37	$-16.77^*$	$-11.49 \pm 0.22$	$-15.43 \pm 0.48$	-1.5	-2.4
0.5	43.31	$-15.89^*$	$-10.80 \pm 0.21$	$-14.82 \pm 0.40$	-1.7	-2.3
0.5	48.70	$-15.27 \pm 0.05$	$-10.03 \pm 0.17$	$-13.97 \pm 0.34$	-1.7	-2.5

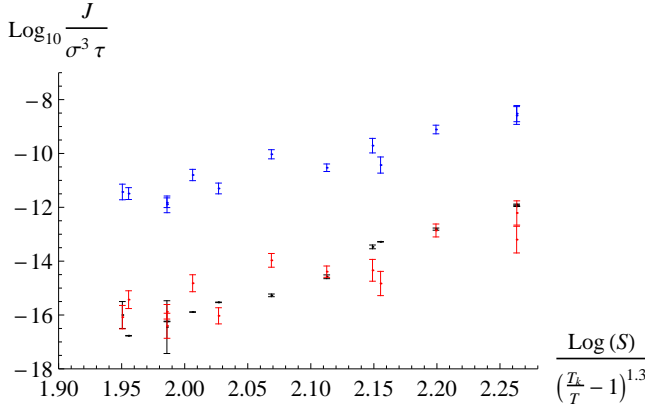


FIG. 6. Comparison of argon nucleation rate predictions of the standard Monte-Carlo technique (blue) with predictions of our theory (red) where the nuclei are described through an extended state space. Also shown are the results of brute-force molecular dynamics simulations<sup>6</sup> (black). Dotted lines are a guide to the eyes.

lowering the ambient temperature  $T$  the nonisothermal effects become more pronounced and comparable with the Brownian effects. In the high temperature regime however, Brownian effects dominate. The lower the temperature the more latent heat is exchanged between the nucleus and ambient phase so that we expect this behavior. In Fig 6 we compare the predictions  $J_{\text{ext}}$  with the brute-force MD rates  $J_{\text{MD}}$  obtained by Diemand et al.<sup>6</sup> and plotted it in a manner that highlights the intrinsic universal scaling behavior suggested by Tanaka et al.<sup>59</sup>,

where the MD nucleation rate is recognized as a universal function of  $\log(S)/(T_k/T - 1)^{1.3}$  with  $T_k = 1.312\epsilon/k_{\text{B}}$  the critical temperature. The rates obtained by the standard truncated theory deviate by 3 – 5 orders of magnitude from the exact MD rates whereas the rates obtained within our extended theory yields a quite excellent agreement.

## VII. CONCLUSION AND OUTLOOK

We have derived the nonequilibrium statistical mechanics of nucleation with an extended state space within the GENERIC coarse-graining technique. We have shown that the truncated state space where only the number of atoms in a nucleus is a relevant variable leads to systematically erroneous prediction of the nucleation rate of 3 – 5 orders of magnitude for the prototypical studied nucleation of liquid drops in a supersaturated Lennard-Jones gas. When we extend the nuclei degree of freedom to involve in addition its internal energy and momentum, this lack in prediction can be significantly improved. For the example of the MFKT of nucleation we also showed that truncated theories can easily be extended to a richer state space to improve predictions. Within our approach we were rigorously able to derive the correction of the nucleation landscape due to energy fluctuations and momentum fluctuations, and in both cases we were able to reproduce the up to date results. This includes the non-Gaussian temperature fluctuations inherent for small nuclei following the law of McGraw



and Laviolette, and the mesoscopic irreversible thermodynamics contribution of Reguera and Rubi for Brownian motion. The latter, as we found, must be corrected by a logarithmic contribution in the cluster size that arises from normalization reasons, and can probably not be obtained by bare thermodynamic argumentations as done by Reguera and Rubi. Within our statistical mechanics approach we were also able to state the proper definition of the nucleus temperature that leads to the temperature fluctuations predicted by McGraw and Laviolette.

We predicted the nucleation rate in a metastable Lennard-Jones gas in the experimentally relevant low supersaturation regime and found excellent agreement with recent large-scale brute-force MD simulations. As argued by Diemand et al. when the nuclei size is smaller than about 16 atoms, nuclei tend to spin so that the angular momentum would enter as an additional important state variable. In the majority of our simulations the critical size was well beyond 16 so that these corrections are irrelevant. From this point of view it might be interesting to include additional state variables into the description such as the angular momentum or the shape of the nucleus. In this respect our coarse-graining procedure could also be extended to multicomponent system for more realistic applications. In our work we relied on a cluster definition to group the atoms as members of the gas phase or the nucleus of interest. Some work has been done avoiding the introduction of an artificial cluster criterion and we also expect useful extensions of our work in this directions, possibly following the lines of Kusaka et al.<sup>26,47</sup>.

### VIII. ACKNOWLEDGMENTS

We thank Hans Christian Öttinger for various hints and ideas to improve this work significantly. Table

- <sup>1</sup>Y. Yang, P. A. Deymier, L. Wang, R. Guzman, and J. B. Hoying, *Biotechnol. Prog.* **22**, 303–312 (2006).
- <sup>2</sup>C. Longuet, A. Yamada, Y. Chen, D. Baigl, and J. Fattaccioli, *J. Chem. Phys.* **386**, 179–182 (2014).
- <sup>3</sup>H. ter Horst, J. H. M. Kramer, and P. J. Jansens, *Cryst. Growth Des.* **2**(5), 351 (2002).
- <sup>4</sup>N. Kuba and M. Murakami, *Atmos. Chem. Phys.* **10**, 33353351 (2010).
- <sup>5</sup>E. Krugel, *The Physics of Interstellar Dust (Series in Astronomy and Astrophysics)* (Institute of Physics Publishing, 2003).
- <sup>6</sup>J. Diemand, R. Angéilil, K. K. Tanaka, and H. Tanaka, *J. Chem. Phys.* **139**, 074309 (2013).
- <sup>7</sup>G. Chkonia, J. Wölk, R. S. Strey, J. Wedekind, and D. Reguera, *J. Chem. Phys.* **130**, 064505 (2009).
- <sup>8</sup>T. Kraska, *J. Chem. Phys.* **124**, 054507 (2006).
- <sup>9</sup>K. K. Tanaka, K. Kawamura, H. Tanaka, and K. Nakazawa, *J. Chem. Phys.* **122**, 184514 (2005).
- <sup>10</sup>K. K. Tanaka, H. Tanaka, T. Yamamoto, and K. Kawamura, *J. Chem. Phys.* **134**, 204313 (2011).
- <sup>11</sup>J. Wedekind, J. Wölk, D. Reguera, and R. Strey, *J. Chem. Phys.* **127**, 154515 (2007).
- <sup>12</sup>K. Yasuoka and M. Matsumoto, *J. Chem. Phys.* **109**, 8451 (1998).
- <sup>13</sup>V. Kalikmanov, *Nucleation Theory* (Springer, 2013).
- <sup>14</sup>S. Girshick and C.-P. Chiu, *J. Chem. Phys.* **93**, 1273 (1990).
- <sup>15</sup>A. Dillmann and G. E. A. Meier, *J. Chem. Phys.* **94**, 3872 (1991).
- <sup>16</sup>C. F. Delale and G. E. A. Meier, *J. Chem. Phys.* **98**, 9850 (1993).
- <sup>17</sup>V. I. Kalikmanov and M. E. H. van Dongen, *J. Chem. Phys.* **103**, 4250 (1995).
- <sup>18</sup>A. Laaksonen, I. Ford, and M. Kulmala, *Phys. Rev. E* **49**, 5517 (1994).
- <sup>19</sup>D. Reguera and H. Reiss, *J. Phys. Chem. B* **108** (51), 19831–19842 (2004).
- <sup>20</sup>J. L. Schmitt, R. A. Zalabsky, and J. L. Adams, *J. Phys. Chem.* **79**, 4496 (1984).
- <sup>21</sup>D. Wright, R. Caldwell, C. Moxely, and M. S. El-Shall, *J. Phys. Chem.* **98**, 3356 (1993).
- <sup>22</sup>Y. Viisanen, R. Strey, and H. Reiss, *J. Phys. Chem.* **99**, 4680 (1993).
- <sup>23</sup>M. P. Anisimov, P. K. Hopke, I. N. Shaimordanov, S. D. Shandakov, and L.-E. Magnusson, *J. Phys. Chem.* **115**, 810 (1993).
- <sup>24</sup>S. Sinha, A. Bhabhe, H. Laksmo, J. Wölk, R. Strey, and B. Wyslouzil, *J. Phys. Chem.* **132**, 064304 (2010).
- <sup>25</sup>V. Kalikmanov, *J. Chem. Phys.* **124**(12), 124505 (2006).
- <sup>26</sup>I. Kusaka, Z.-G. Wang, and J. H. Seinfeld, *J. Phys. Chem.* **108**, 3416 (1998).
- <sup>27</sup>K. J. Oh and X. C. Zeng, *J. Phys. Chem.* **110**, 4471 (1999).
- <sup>28</sup>K. J. Oh and X. C. Zeng, *J. Phys. Chem.* **112**, 294 (2000).
- <sup>29</sup>B. Chen, I. Siepmann, J. O. Kwang, and M. L. Klein, *J. Phys. Chem.* **115**, 10903 (2001).
- <sup>30</sup>S. Yoo, K. J. Oh, and X. C. Zeng, *J. Phys. Chem.* **115**, 8518 (2001).
- <sup>31</sup>M. A. Gonzalez, E. Sanz, C. McBride, J. L. F. Abascal, C. Vegaa, and C. Valeriani, *J. Phys. Chem.* **16**, 24913–24919 (2014).
- <sup>32</sup>P. R. ten Wolde and D. Frenkel, *J. Phys. Chem.* **109**, 9901 (1998).
- <sup>33</sup>M. Schweizer and L. Sagis, *J. Chem. Phys.* **141**, 224102 (2014).
- <sup>34</sup>J. Feder, J. Lothe, and G. M. Pound, *Adv. Phys.* **15**, 111 (1966).
- <sup>35</sup>J. C. Barrett, *J. Chem. Phys.* **128**, 164519 (2008).
- <sup>36</sup>J. C. Barrett, *J. Phys. A: Math. Gen.* **27**, 5053 (1994).
- <sup>37</sup>B. E. Wyslouzil and J. H. Seinfeld, *J. Chem. Phys.* **97**, 2661 (1992).
- <sup>38</sup>J. Lothe and G. M. Pound, *J. Chem. Phys.* **36**, 2080 (1962).
- <sup>39</sup>H. Reiss, W. K. Kegel, and J. L. Katz, *Phys. Chem. A* **102**, 8548 (1998).
- <sup>40</sup>H. Reiss, W. K. Kegel, and J. L. Katz, *Phys. Rev. Lett.* **78**, 4506 (1997).
- <sup>41</sup>V. Ruth, J. P. Hirth, and G. M. Pound, *J. Chem. Phys.* **88**, 7079 (1988).
- <sup>42</sup>H. Reiss, J. L. Katz, and E. R. Cohe, *J. Chem. Phys.* **48**, 5553 (1968).
- <sup>43</sup>H. Reiss and J. L. Katz, *J. Chem. Phys.* **46**, 2496 (1967).
- <sup>44</sup>D. Reguera and J. M. Rubi, *J. Chem. Phys.* **115**, 7100 (2001).
- <sup>45</sup>P. Schaaf, B. Senger, J.-C. Voegel, and H. Reiss, *Phys. Rev. E* **60**, 771 (1999).
- <sup>46</sup>I. Kusaka, *J. Chem. Phys.* **119**, 3820 (2003).
- <sup>47</sup>I. Kusaka and D. W. Oxtoby, *J. Chem. Phys.* **110**, 5249 (1999).
- <sup>48</sup>H. C. Öttinger, *MRS Bull.* **32**, 936–940 (2007).
- <sup>49</sup>H. C. Öttinger, *Beyond Equilibrium Thermodynamics* (Wiley Interscience, 2005).
- <sup>50</sup>S. E. M. Lundrigan and I. Saika-Voivod, *J. Chem. Phys.* **131**, 104503 (2009).
- <sup>51</sup>S. Auer and D. Frenkel, *J. Chem. Phys.* **120**, 3015 (2004).
- <sup>52</sup>R. McGraw and L. R. A., *J. Chem. Phys.* **102**, 8983 (1995).
- <sup>53</sup>J. Wedekind, R. D., and R. Strey, *J. Chem. Phys.* **127**, 064501 (2007).
- <sup>54</sup>R. Angéilil, J. Diemand, K. K. Tanaka, and H. Tanaka, *J. Chem. Phys.* **140**, 074303 (2014).
- <sup>55</sup>J. Wedekind and D. Reguera, *J. Chem. Phys.* **127**(15), 154516 (2007).
- <sup>56</sup>D. Reguera and J. M. Rubi, *J. Chem. Phys.* **119**, 9888 (2003).
- <sup>57</sup>D. Reguera and J. M. Rubi, *J. Chem. Phys.* **119**, 9877 (2003).
- <sup>58</sup>H. Trinkaus, *Phys. Rev. B* **27**, 7372 (1983).
- <sup>59</sup>K. K. Tanaka, J. Diemand, R. Angéilil, and H. Tanaka, *J. Chem. Phys.* **140**, 194310 (2014).

## IX. APPENDIX

### A. Truncated Entropy Splitting Eq. (10)

Without restriction of generality, we can assume  $P_{\text{tot}} = 0$  emphasizing Galilean invariance. When the environment is sufficiently large, the variables  $\mathbf{x}_{\text{tot}}$  reach their thermodynamic limit and by the equivalence of ensembles we then have  $S(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) = \log(Z(\mathbf{x}_{\text{tot}}, \mathbf{x}_c))$ , where

$$\begin{aligned} Z(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) &= \int_{V_{\text{tot}}} d\Gamma \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c) e^{-\frac{\Pi_{E_{\text{tot}}}}{2T}} \\ &= e^{-\frac{p_c^2}{2n_c T}} Z(\mathbf{x}_{\text{tot}}, n_c, \mathbf{p}_c = 0, n_c). \end{aligned} \quad (46)$$

The second line follows by straight forward calculation. This can be related to the truncated partition function  $Z(\mathbf{x}_{\text{tot}}, n_c, u_c)$  which finally defines the truncated entropy  $S(\mathbf{x}_{\text{tot}}, n_c, u_c)$  by means of

$$\begin{aligned} Z(\mathbf{x}_{\text{tot}}, n_c, u_c) &= \int d^3\mathbf{p}_c Z(\mathbf{x}_{\text{tot}}, \mathbf{x}_c) \\ &= \left(\frac{n_c}{T}\right)^{3/2} Z(\mathbf{x}_{\text{tot}}, n_c, \mathbf{p}_c = 0, n_c) \\ &= \left(\frac{n_c}{T}\right)^{3/2} e^{\frac{p_c^2}{2n_c T}} Z(\mathbf{x}_{\text{tot}}, \mathbf{x}_c). \end{aligned} \quad (47)$$

The required relationship (10) then follows by exploiting Galilean invariance again to shift to the general case of arbitrary  $P_{\text{tot}}$ .

### B. Detailed Derivation of Evolution Equation

We use the GENERIC projection operator technique<sup>48,49</sup> in order to find the fundamental time evolution of the distribution function  $f$  of the nucleus of interest. To this extend a projection operator  $\mathcal{P}$  is introduced that separates the macroscopically relevant contribution of an arbitrary observable  $\Pi_A$  on the phase space  $\bar{\Gamma}$  from the irrelevant ones to be eliminated - this corresponds to the separation of slow, relevant from fast, irrelevant degrees of freedom when switching from the finer-level of description involving all  $N$  atoms to the coarse-grained level characterized by  $(\mathbf{x}_{\text{tot}}, f)$ . Following the standard definition, we define the projection onto the slow variables

$$\begin{aligned} \mathcal{P}\Pi_A &= \langle \Pi_A \rangle_{\rho(\mathbf{x}_{\text{tot}}, \lambda_c)} \\ &+ (\Pi_{\mathbf{x}_{\text{tot}}} - \mathbf{x}_{\text{tot}}) \cdot \frac{\delta}{\delta \mathbf{x}_{\text{tot}}} \langle \Pi_A \rangle_{\rho(\mathbf{x}_{\text{tot}}, \lambda_c)} \\ &+ \int d\mathbf{x}_c (\Pi_{f(\mathbf{x}_c)} - f(\mathbf{x}_c)) \frac{\delta}{\delta f(\mathbf{x}_c)} \langle \Pi_A \rangle_{\rho(\mathbf{x}_{\text{tot}}, \lambda_c)}. \end{aligned} \quad (48)$$

Using

$$\rho_{\mathbf{x}_{\text{tot}}, \lambda_c} \Pi_{f(\mathbf{x}_c)} = f(\mathbf{x}_c) \rho_{\mathbf{x}_{\text{tot}}, \mathbf{x}_c}, \quad (49)$$

where

$$\rho_{\mathbf{x}_{\text{tot}}, \mathbf{x}_c} \propto \delta(\Pi_{\mathbf{x}_{\text{tot}}} - \mathbf{x}_{\text{tot}}) \delta(\Pi_{\mathbf{x}_c} - \mathbf{x}_c), \quad (50)$$

is the fully microcanonical ensemble, the projection operator can be expressed alternatively as

$$\begin{aligned} \mathcal{P}\Pi_A &= \langle \Pi_A \rangle_{(\mathbf{x}_{\text{tot}}, \mathbf{x}_c = \Pi_{\mathbf{x}_c})} \\ &+ (\Pi_{\mathbf{x}_{\text{tot}}} - \mathbf{x}_{\text{tot}}) \cdot \frac{\delta}{\delta \mathbf{x}_{\text{tot}}} \int d\mathbf{x}_c f(\mathbf{x}_c) \langle \Pi_A \rangle_{(\mathbf{x}_{\text{tot}}, \mathbf{x}_c)}. \end{aligned} \quad (51)$$

The first part projects onto the nucleus variables  $\Pi_{\mathbf{x}_c}$  while the second part onto the system state  $\mathbf{x}_{\text{tot}}$ . Complementary, we define  $\mathcal{Q} = 1 - \mathcal{P}$  as the projection onto the fast variables. Denoting by  $i\mathcal{L}$  the Liouville operator on  $\bar{\Gamma}$ , we notice that  $i\mathcal{L}\Pi_{\mathbf{x}_{\text{tot}}} = 0$ . The fact that  $\mathbf{x}_{\text{tot}}$  are conserved quantities also implies on the macroscale  $d\mathbf{x}_{\text{tot}}/dt = 0$ . The projection operator technique then implies<sup>48,49</sup> the time-evolution for  $f$ ,

$$\begin{aligned} \frac{\partial f(\mathbf{x}_c)}{\partial t} &= \langle i\mathcal{L}\Pi_{f(\mathbf{x}_c)} \rangle_{\rho(\mathbf{x}_{\text{tot}}, \lambda_c)} \\ &+ \int d\mathbf{x}'_c M_{f(\mathbf{x}_c), f(\mathbf{x}'_c)} \frac{\delta S(\mathbf{x}_{\text{tot}}, f)}{\delta f(\mathbf{x}'_c)}, \end{aligned} \quad (52)$$

where

$$M_{f(\mathbf{x}_c), f(\mathbf{x}'_c)} = \frac{1}{k_B} \int_0^{\tau_{\text{GK}}} du \langle \dot{\Pi}_{f(\mathbf{x}'_c)} G(u) \dot{\Pi}_{f(\mathbf{x}_c)} \rangle_{\rho(\mathbf{x}_{\text{tot}}, \lambda_c)}. \quad (53)$$

Here,  $\dot{\Pi}_{f(\mathbf{x}_c)} = \mathcal{Q}i\mathcal{L}\Pi_{f(\mathbf{x}_c)}$  is the fast time-evolution of  $\Pi_f$  and  $G(u) = e^{\mathcal{Q}i\mathcal{L}\mathcal{Q}u}$  is the time-evolution operator. the time evolution equation for  $f$ , Eq. (52), splits into a reversible part and an irreversible part involving the Green-Kubo coefficients  $M_{f(\mathbf{x}_c), f(\mathbf{x}'_c)}$  and driven by the total entropy  $S(\mathbf{x}_{\text{tot}}, f)$  of the system. The time-scale  $\tau_{\text{GK}}$  is an intermediate time-scale between microscopic collisions and the mesoscopic time-scale on which the nucleation process is described in a coarse-grained way.

Using the auxiliary properties

$$i\mathcal{L}\Pi_{f(\mathbf{x}_c)} = -\frac{\delta \Pi_{f(\mathbf{x}_c)}}{\delta \mathbf{x}_c} \cdot i\mathcal{L}\Pi_{\mathbf{x}_c}, \quad (54)$$

and (50) the reversible part of the time-evolution can be cast into

$$\begin{aligned} \left. \frac{\partial f(\mathbf{x}_c)}{\partial t} \right|_{\text{rev}} &= \langle i\mathcal{L}\Pi_{f(\mathbf{x}_c)} \rangle_{\rho(\mathbf{x}_{\text{tot}}, \lambda_c)} \\ &= -\frac{\delta}{\delta \mathbf{x}_c} f(\mathbf{x}_c) \langle i\mathcal{L}\Pi_{\mathbf{x}_c} \rangle_{\rho_{\mathbf{x}_{\text{tot}}, \mathbf{x}_c}}. \end{aligned} \quad (55)$$

We now turn our attention back to the projection operator  $\mathcal{P}$ . This operator, or its counterpart  $\mathcal{Q}$  act on  $i\mathcal{L}\Pi_{\mathbf{x}_c}$  in the evolution equation. Therefore, we must understand the object  $\langle i\mathcal{L}\Pi_{\mathbf{x}_c} \rangle_{\rho_{\mathbf{x}_{\text{tot}}, \mathbf{x}_c}}$  which also enters the

reversible part of the evolution equation. For homogeneous nucleation this reversible part is known to vanish, so that we already expect

$$\langle i\mathcal{L}\Pi_{\mathbf{x}_c} \rangle_{\rho_{\mathbf{x}_{\text{tot}}, \mathbf{x}_c}} = 0 \quad (56)$$

This implies  $\mathcal{P}i\mathcal{L}\Pi_{\mathbf{x}_c} = 0$  and  $\mathcal{Q}i\mathcal{L}\Pi_{\mathbf{x}_c} = i\mathcal{L}\Pi_{\mathbf{x}_c}$ . The argument for Eq. (56) can be made more formal by recognizing that the average over  $i\mathcal{L}\Pi_{\mathbf{x}_c}$  must be taken over all states compatible with the microcanonical ensemble  $\rho_{\mathbf{x}_{\text{tot}}, \mathbf{x}_c}$ , that is over all states for which we encounter a nucleus with properties  $\mathbf{x}_c$ . But if a certain atomic configuration  $C = (\mathbf{r}_1, \dots, \mathbf{r}_N, \mathbf{p}_1, \dots, \mathbf{p}_N)$  contains a nucleus with properties  $\mathbf{x}_c$ , then also the mirror image  $C' = (-\mathbf{r}_1, \dots, -\mathbf{r}_N, \mathbf{p}_1, \dots, \mathbf{p}_N)$  obeys this requirement. It is important that the momentum variables must not be mirror imaged, since otherwise the momentum  $\Pi_{\mathbf{p}_c}$  of  $C$  and  $C'$  do not match and therefore  $C'$  would not fall into the class  $\mathbf{x}_c$ . In  $C$  all particles that would be about to join the nucleus would be about to leave it in  $C'$  and vice versa. In the same manner one can argue that in

$C$  a momentum influx is transformed to a momentum outflux in  $C'$  and vice versa. The same argumentation holds with the energy flux and consequently we have  $i\mathcal{L}\Pi_{\mathbf{x}_c}(C) = -i\mathcal{L}\Pi_{\mathbf{x}_c}(C')$  and therefore the validity of Eq. (56).

In general, the irreversible part of the evolution equation can be cast into the form<sup>48,49</sup>

$$\begin{aligned} \left. \frac{\partial f(\mathbf{x}_c)}{\partial t} \right|_{\text{irr}} &= \int d\mathbf{x}'_c M_{f(\mathbf{x}_c), f(\mathbf{x}'_c)} \frac{\delta S(\mathbf{x}_{\text{tot}}, f)}{\delta f(\mathbf{x}'_c)} \\ &= -\frac{\delta}{\delta \mathbf{x}_c} \cdot f(\mathbf{x}_c) M(\mathbf{x}_c) \cdot \frac{\delta S(\mathbf{x}_{\text{tot}}, f)}{\delta \mathbf{x}_c}, \end{aligned} \quad (57)$$

where

$$\begin{aligned} M(\mathbf{x}_c) &= \frac{1}{k_B} \int_0^{\tau_{\text{GK}}} du \langle i\mathcal{L}\Pi_{\mathbf{x}_c} G(u) i\mathcal{L}\Pi_{\mathbf{x}_c} \rangle_{\rho(\mathbf{x}_{\text{tot}}, \mathbf{x}_{\text{tot}})} \\ &= \frac{1}{2k_B \tau_{\text{GK}}} \langle \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \Delta_{\tau_{\text{GK}}} \Pi_{\mathbf{x}_c} \rangle_{\rho(\mathbf{x}_{\text{tot}}, \mathbf{x}_{\text{tot}})}. \end{aligned} \quad (58)$$

We hence derived the fundamental nucleus evolution equation based on microscopic considerations.